

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
17 January 2002 (17.01.2002)

PCT

(10) International Publication Number
WO 02/04488 A2

(51) International Patent Classification⁷: C07K 14/00

(21) International Application Number: PCT/GB01/03130

(22) International Filing Date: 12 July 2001 (12.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/614,679 12 July 2000 (12.07.2000) US

(71) Applicants (for all designated States except US): SANG-AMO BIOSCIENCES, INC. [US/US]; Point Richmond Tech Center, Suite A100, 501 Canal Boulevard, Richmond, CA 94804 (US). CAMBRIDGE UNIVERSITY TECHNICAL SERVICES LTD. [GB/GB]; The Old Schools, Trinity Lane, Cambridge CB2 1TN (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): CHOO, Yen [GR/GB]; 130 Ross Street, Cambridge CB1 3BU (GB).

ISALAN, Mark [GB/GB]; 24 Shottfield Avenue, East Sheen, London SW14 8EA (GB). LIU, Xiaohai [CN/GB]; 136 Mowbray Road, Cambridge, CB1 7TG (GB). PATEL, Sachin [IN/IN]; Themis Medicare, 10/12 Udyog Nagar, S-V-RD, Goregon (W), Mumbai 400104 (IN). BALASUBRAMANIAN, Shankar [GB/GB]; Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW (GB).

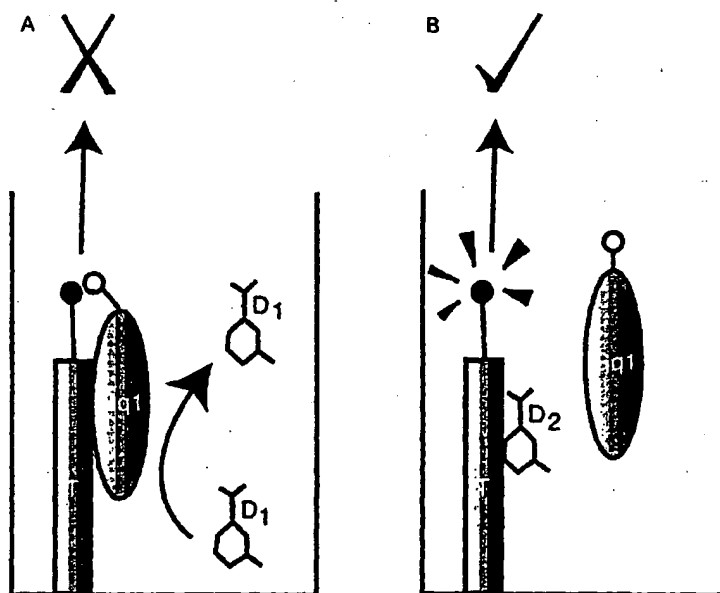
(74) Agents: KHOO, Chong-Yee et al.; D Young & Co., 21 New Fetter Lane, London EC4A 1DA (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian

[Continued on next page]

(54) Title: NUCLEIC ACID BINDING POLYPEPTIDES



(57) Abstract: We describe the use of a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid as an inhibitor of enzymatic activity, including a telomerase activity, a polymerase activity, an integrase activity and a gp120 activity. Telomerase assays as well as methods of identifying molecules capable of interacting with telomeric, G-quadruplex, or G-quartet nucleic acid are also described.



patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

Declaration under Rule 4.17:

— of inventorship (Rule 4.17(iv)) for US only

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

NUCLEIC ACID BINDING POLYPEPTIDES

FIELD OF THE INVENTION

The invention relates to nucleic acid binding polypeptides. In particular the invention relates to molecules which bind to G-quadruplex or telomeric DNA.

5 BACKGROUND TO THE INVENTION

There is considerable interest in molecules that bind to telomeric DNA sequences and G-quadruplexes. Such molecules will be useful to test hypotheses of telomere length regulation, and may have therapeutic potential.

Several naturally occurring proteins with affinity for G-quadruplexes have been
10 described in the prior art (reviewed in Wellinger, R. J., & Sen, D. (1997) *European Journal of Cancer* 33, 735-749), although none have so far proved to be good candidates for use as diagnostic probes or therapeutic tools.

Prior art quadruplex DNA binding molecules, such as a recently reported DNA-binding autoantibody (Brown, B. A., Li, Y. Q., Brown, J. C., Hardin, C. C., Roberts, J. F.,
15 Pelsue, S. C., & Shultz, L. D. (1998) *Biochemistry* 37, 16325-16337), have only moderate binding affinities and discriminate weakly between duplex and quadruplex DNA.

DNA binding molecules are disclosed in M. D. Isalan, A. Klug and Y. Choo, International Patent Application Publication No. WO98/53057.

Naturally occurring telomere-binding proteins are also unable to discriminate these
20 structures. For example, *Saccharomyces cerevisiae* RAP1 (Giraldo, R., & Rhodes, D. (1994) *EMBO J* 13, 2411-2420) has distinct but inseparable domains for binding quadruplexes and double stranded DNA.

The present invention seeks to overcome problems associated with the prior art.

SUMMARY OF THE INVENTION

Thus, in a first aspect, the invention relates to use of a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid as an inhibitor of enzymatic activity.

- 5 There is provided, according to a second aspect of the present invention, a method of inhibiting an enzymatic activity, the method comprising: (a) providing an enzyme; and (b) contacting the enzyme with a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid.

- 10 Preferably, the use or method further comprises the step of providing a telomeric, G-quadruplex, or G-quartet nucleic acid and contacting the nucleic acid with the enzyme and/or the nucleic acid binding polypeptide. More preferably, the enzymatic activity is selected from the group consisting of: a telomerase activity, a polymerase activity, an integrase activity and a gp120 activity. Most preferably, the enzymatic activity is inhibited *in vivo*.

- 15 We provide, according to a third aspect of the present invention, a method of preventing replication of a retrovirus, the method comprising exposing the retrovirus or a nucleic acid portion thereof to a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid. Preferably, the retrovirus is Human Immunodeficiency Virus.

- 20 As a fourth aspect of the present invention, there is provided a method of treatment of a patient suffering from a disease, the method comprising administering to a patient in need of such treatment a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid.

- 25 Preferably, the disease comprises infection by Human Immunodeficiency Virus infection. Preferably, the disease comprises a hyperproliferative disease, preferably cancer.

We provide, according to a fifth aspect of the present invention, a method for assaying telomerase activity, the method comprising: (i) providing a nucleic acid substrate for telomerase; (ii) contacting the nucleic acid substrate with a telomerase; (iii) contacting the nucleic acid substrate with a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid; and (iv) monitoring the binding of the nucleic acid binding polypeptide to the nucleic acid substrate.

The present invention, in a sixth aspect, provides a method for determining the length of a telomere, the method comprising: (i) contacting the telomere with a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid; (ii) monitoring the binding of the nucleic acid binding polypeptide to the telomere, and (iii) determining the length of the telomeres from the strength of the binding.

In a seventh aspect of the present invention, there is provided a method for discriminating between duplex and quadruplex nucleic acid comprising contacting a sample of nucleic acid with a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid, and monitoring the binding of the nucleic acid binding polypeptide to the nucleic acid.

According to an eighth aspect of the present invention, we provide a method of detecting telomeric structures in a system, the method comprising: (a) exposing the system to a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid; (b) detecting binding between the nucleic acid binding polypeptide and any telomeric structures in the system.

Preferably, the nucleic acid binding polypeptide is labelled. More preferably, the location of binding is detected to localise telomeric structures in the system. Most preferably the system comprises a cell and binding is detected *in vivo* or *in situ*.

We provide, according to a ninth aspect of the invention, a method of identifying a molecule capable of binding to a telomeric, G-quadruplex, or G-quartet structure in a nucleic acid, the method comprising: (a) providing a nucleic acid comprising a telomeric, G-quadruplex, or G-quartet structure; (b) providing a nucleic acid binding polypeptide capable of binding to a nucleic acid comprising such a structure; (c) contacting either or both of the nucleic acid and the nucleic acid binding polypeptide with a candidate molecule; and (d) determining the binding between the nucleic acid and the nucleic acid binding polypeptide.

There is provided, in accordance with a tenth aspect of the present invention, a method of identifying a molecule capable of binding to a telomeric, G-quadruplex, or G-quartet structure in a nucleic acid, the method comprising monitoring the binding between a nucleic acid comprising a telomeric, G-quadruplex, or G-quartet structure and a nucleic acid binding polypeptide capable of binding to a nucleic acid comprising such a structure, in the presence and absence of a candidate molecule.

As an eleventh aspect of the invention, we provide a method of identifying a molecule capable of binding to a telomeric, G-quadruplex, or G-quartet structure in a nucleic acid, the method comprising providing a complex between a nucleic acid comprising a telomeric, G-quadruplex, or G-quartet structure and a nucleic acid binding polypeptide capable of binding to a nucleic acid comprising such a structure; contacting either or both members of the complex with a candidate molecule; and detecting a dissociation between the members of the complex.

Preferably, the candidate molecule is provided in the form of a library of candidate molecules, more preferably an array of candidate molecules. The method may further comprise a step of isolating, synthesising and/or providing a composition comprising the candidate molecule identified to have such activity.

The binding or dissociation between the nucleic acid binding polypeptide and the nucleic acid may be monitored by various means. In a preferred embodiment, the

monitoring is by means of an ELISA assay. Alternatively or in addition, the binding or dissociation may be monitored by detecting Fluorescence Resonance Energy Transfer (FRET).

The binding or dissociation is preferably monitored in a micro-well.

5 We provide, according to a twelfth aspect of the invention, a method for manipulating telomeric structure(s) *in vivo* comprising contacting a labelled nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid with a telomeric structure, in which the nucleic acid binding polypeptide further comprises an effector domain.

10 According to a thirteenth aspect of the present invention, we provide a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid for use in a method of treatment of a disease.

There is provided, according to a fourteenth aspect of the present invention, use of a nucleic acid binding polypeptide capable of binding to one or more of telomeric;
15 G-quadruplex, or G-quartet nucleic acid for the preparation of a pharmaceutical composition for the treatment of a disease.

Preferably, the disease comprises a retroviral infection, infection with Human Immunodeficiency Virus, or AIDS.

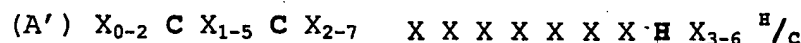
Preferably, the nucleic acid is not in a double-helical conformation. More
20 preferably, the nucleic acid comprises single-stranded DNA. Most preferably, the nucleic acid is comprised in a chromosome end. In a highly preferred embodiment, the nucleic acid is comprised in a telomeric structure. The nucleic acid may be in a non-Watson-Crick base paired conformation, preferably comprising Hoogsteen base pairing. Preferably, the nucleic acid binding polypeptide has an affinity for G-quadruplex nucleic acid which is
25 different from its affinity for duplex nucleic acid. A preferred nucleic acid binding

polypeptide is one which binds to any one or more of the nucleic acids having the preferred properties set out above.

- In a highly preferred embodiment of the invention, the nucleic acid binding polypeptide comprises a zinc finger motif. Most preferably, a nucleic acid binding polypeptide or zinc finger comprises any of the following structures:

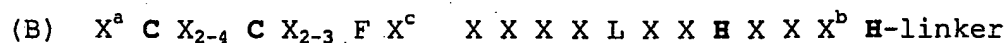


where X is any amino acid, and the numbers in subscript indicate the possible numbers of residues represented by X;



-1 1 2 3 4 5 6 7

- where X is any amino acid, and the numbers in subscript indicate the possible numbers of residues represented by X; or



-1 1 2 3 4 5 6 7 8 9

where X (including X^a , X^b and X^c) is any amino acid. X_{2-4} and X_{2-3} refer to the presence of 2 or 4, or 2 or 3, amino acids, respectively.

- Preferred embodiments of the invention utilise nucleic acid binding polypeptides and/or zinc fingers in which the amino acids at positions -1, 1, 2, 3, 4, 5 and 6 are selected from the group consisting of: RDS AHLTR, DRSDLSE, RSDHRIE, RSDHLIN, DRADLSE, TSSHRTN, DSAHLTR, DRDHLSE, TSSHRTN, TSHHLIQ, DRADLSE, and HQHYRTN. More preferably, the polypeptide comprises three zinc finger motifs F1,

F2 and F3, in which the amino acids at positions -1, 1, 2, 3, 4, 5 and 6 of F1, F2 and F3 comprise: F1: DSAHLTR, F2: DRSDLSE, F3: RSDHRIE. Most preferably, the nucleic acid binding polypeptide comprises a sequence derived from at least one of the fingers of Gq1.

5 We provide, according to a fifteenth aspect of the present invention, use of a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid as a cytotoxic agent.

As a sixteenth aspect of the present invention, there is provided a method of killing a cell, which method comprises exposing a cell to a nucleic acid binding polypeptide
10 capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid.

We provide, according to a seventeenth aspect of the present invention, a nucleic acid binding polypeptide comprising a sequence selected from the group consisting of: Gq1(1:3)-linkerA-Gq1(1:3) amino acid sequence, Gq1(1:3)-linkerB-Gq1(1:3) amino acid sequence, Gq1(1:2)-linkerA-Gq1(1:2) amino acid sequence, Gq1(1:2)-linkerB-Gq1(1:2)
15 amino acid sequence, and fragments or derivatives of the above.

According to an eighteenth aspect of the present invention, we provide a nucleic acid sequence capable of encoding a nucleic acid binding polypeptide according to the seventeenth aspect of the invention.

Preferably, the nucleic acid sequence is selected from the group consisting of:
20 Gq1(1:3)-linkerA-Gq1(1:3) nucleic acid sequence, Gq1(1:3)-linkerB-Gq1(1:3) nucleic acid sequence, Gq1(1:2)-linkerA-Gq1(1:2) nucleic acid sequence, Gq1(1:2)-linkerB-Gq1(1:2) nucleic acid sequence, and fragments or derivatives of the above.

We provide, according to a nineteenth aspect of the invention, a use, method or a nucleic acid binding polypeptide according to any preceding aspect, in which the nucleic
25 acid binding polypeptide comprises a polypeptide according to the seventeenth aspect of

the invention, or a polypeptide encoded by a nucleic acid sequence according to the eighteenth aspect of the invention.

BRIEF DESCRIPTION OF THE FIGURES

The invention will now be described by way of example, with reference to the
5 following figures:

Figure 1A shows schematic representation of the Zif268 DNA-binding domain, indicating its three zinc finger helices (F1, F2 and F3). The circled numbers represent the key amino acid residues that interact with duplex DNA (relative to the first position of the α -helix, position +1).

10 Figure 1B shows amino acids included in the phage display library used in this study. Amino acid residues in the helical regions of fingers 1-3 (F1-F3) are shown in single letter code, numbered relative to the first helical position (position +1). Note that library construction involved cloning a subset of the possible combinations shown above, although these clones are pre-enriched for DNA-binding potential (See below).

15 Figure 2A shows DMS methylation protection analysis of Htelo. End-labelled ^{32}P -Htelo is annealed in KCl or NaCl at the indicated concentrations. Each sample is incubated with DMS for 5 minutes and then cleaved with piperidine. Methylation protection patterns, indicative of G-quadruplex formation, appear after resolution of the cleaved fragments on a 20% polyacrylamide gel. The Tris control lane indicates the reference
20 (non-quadruplex) methylation cleavage pattern of Htelo in the absence of Na^+ or K^+ .

Figure 2B shows schematic representation of an exemplary isoform of an intramolecular antiparallel G-quadruplex formed by Htelo. Guanines in the G-quartet core are labelled in shaded circles with darker shading indicating a relatively higher amount of cleavage, as observed in the DMS methylation protection analysis. (Note that the structure
25 shown is only one possible isoform and that other 'semi-parallel' conformation(s) such as

one comprising a pair of parallel 'up' strands, facing a pair of parallel 'down' strands created by 'crossing-over' of the two top 'TTA' sequences in the figure may also be stable form(s) of Htelo.)

Figure 3 shows peptide sequences of the zinc finger helical domains of the four proteins Gq1-4, obtained after three rounds of selection. Amino acid residues in fingers 1-3 (F1-F3) are shown in single letter code, numbered relative to the first helical position (position +1). The zinc finger helices of the wild-type Zif268 DNA-binding domain are also shown for comparison.

Figure 4 shows apparent equilibrium binding curves for protein Gq1 binding to single-stranded DNA sequences, and to the Htelo duplex sequence, as measured by phage ELISA. All ELISA procedures are carried out in the presence of 150 mM K⁺, to stabilise G-quadruplexes.

Figure 5A shows gel mobility shift assay of Gq1* binding to Htelo. The analysis is carried out in 8% non-denaturing polyacrylamide gel at 4°C. The DNA concentration is fixed at 1 nM while the amount of protein added to the binding reaction is varied as follows: 800 nM (lane 1), 400 nM (lane 2), 200 nM (lane 3), 100 nM (lane 4), 50 nM (lane 5), 25 nM (lane 6), 12.5 nM (lane 7) and 0 nM (lane 8)

Figure 5B shows apparent equilibrium binding curve obtained by calculating the fraction of Htelo bound at varying Gq1* concentrations (Imagequant software). The binding constant is determined by fitting to the equation $\theta = [P] / \{K_d + [P]\}$ as described in the Examples.

Figure 6 shows DMS methylation protection analysis of Htelo in the presence of Gq1* protein. Htelo is annealed in 100 mM K⁺ or 50 mM Tris-HCl, and methylation protection patterns are obtained in the presence or absence of 200 nM Gq1* (ie. 200 nM Gq1* - a concentration giving approximately full shift). DNA concentration is 1nM. Each sample is incubated with DMS for 5 min. Fragments formed by piperidine cleavage of

5 methylated guanines, are resolved on a 20% polyacrylamide gel. Lane 1: methylation pattern of Htelo in the presence of 100 mM K^+ ; Lane 2: reference methylation pattern of Htelo in the absence of K^+ ; Lane 3: methylation pattern of Htelo in the presence of 100 mM K^+ and incubation with 200 nM Gq1*; Lane 4: methylation pattern in the absence of K^+ , incubated with 200 nM Gq1*.

Figure 7 shows Table 2 which shows apparent ELISA dissociation constants (K_d^E) of the phage-displayed zinc finger peptide, Gq1, for variants of the Htelo DNA sequence. ELISAs from which binding is too low to determine K_d are denoted by a dash (-).

10 Figure 8 shows a sensorgram of various DNA sequences binding to Gq1-GST, as assayed by surface plasmon resonance. The sensorgrams are used to attain the binding constants in the table using equations described in Example 4. A corresponding table is shown as Table 3.

Figure 9 shows a schematic illustration of the "DNA polymerase stop assay". A 13-mer oligonucleotide is used to prime a 50-mer template (Htemp), using the Klenow
15 Fragment of *E. coli* DNA polymerase I. The 50-mer template is designed such that it contains a 24 nucleotide telomeric region 5'-(TTA GGG)₄-3' that can fold into an intramolecular G-quadruplex. The 13-mer primer may be extended by Klenow fragment to form full-length 50-mer product. Alternatively, the G-quadruplex structure may result in a pause site (23-mer) in the extension reaction. This assay is used to evaluate whether the
20 stability of the G-quadruplex structure can be altered by the binding of an engineered zinc finger protein, Gq1.

Figure 10A shows a gel mobility shift assay for Gq1 binding to the Htemp 50-mer DNA template. The DNA concentration is fixed at 1 nM while the concentration of Gq1 protein added to the binding reaction is varied as shown above each lane. Binding is
25 carried out in 100 mM K^+ to promote G-quadruplex formation. Figure 10B shows an equilibrium binding curve obtained by calculating the fraction of Htemp bound at varying

Gq1 concentrations (ImageQuant software). The binding constant (K_d) is determined by fitting to the equation $\theta = [P] / \{K_d + [P]\}$ (See Materials and Methods).

Figure 11 shows DMS methylation protection analysis of Htemp DNA in the presence of Gq1 protein. Htemp (1 nM) is annealed in either 100 mM K^+ [to promote G-quadruplex formation] or in 20 mM Tris-HCl [to destabilise quadruplex structures].
5 Methylation protection patterns are then obtained in either the presence or absence of excess Gq1 (500 nM). Each sample is incubated with DMS for 5 minutes, and the fragments formed by piperidine cleavage of methylated guanines are resolved on a 20 % polyacrylamide gel.

10 Figures 12A and 12B show a DNA polymerase stop assay. Primer extension reactions are carried out with the Klenow (Exo⁻) fragment on the Htemp template [as shown schematically in Figure 9]. Figure 12A is a gel showing enhanced pausing of DNA synthesis at the G-quadruplex site with increasing concentration of Gq1 (Lanes 1-5). The 50mer band indicates the full-length product of DNA synthesis, while the 23mer band is a
15 result of the pause site that is immediately 3' to the G-quadruplex structure. A 13-mer band is present due to residual unextended primer. Figure 12B shows quantitation of the gel using ImageQuant software. The intensity of the paused (23-mer) bands are normalised as a fraction of the total radioactive intensity in each lane, and plotted against the concentration of Gq1 protein in each stop assay.

20 Figure 13 shows the effect of Gq1 on the inhibition of telomerase activity studied by a modified TRAPEZE assay. Telomerase/Gq1 reactions are treated to remove all proteins prior to PCR detection of telomerase extension products. Lanes 1 - 6: the activity of telomerase in the presence of Gq1 concentrations ranging from 0 to 375 nM. Lane 7: control where the telomerase extract is heat-inactivated (90 °C, 10 min). Lanes 8 and 9:
25 PCR amplification of an 8-telomeric-repeat control (TSR8; not treated with telomerase extract) in the presence or absence of a large excess of Gq1 (2.5 mM). Lane 10: internal PCR control experiment.

Figure 14 shows a plot of the quantitated telomerase activity in each lane against Gq1 concentration. The IC₅₀ value is calculated by fitting the data to the equation $y = 100 / (1 + (I / IC_{50}))$.

Figure 15 shows the effect of Gq1 on PCR amplification of telomeric DNA. Gq1 at varying concentration is pre-incubated with or without telomerase for 10 min at ambient temperature prior to initiating the telomerase reaction by addition of dNTP's, TS primer, Taq polymerase, and the PCR primers [PCR mix 1 containing RP + ICT + NT]. Control experiments are also carried out at various concentration of Gq1 where instead of telomerase, a TSR8 template containing 8 telomeric repeats is added. All the reaction mixtures are incubated for 30 min at 30 °C, after which the samples are PCR amplified (two-step cycle of 30 s at 94 °C, 30 s at 59 °C for 30 cycles). Amplified telomerase products are resolved by PAGE and quantitated by a phosphorimager. The concentration of Gq1 is increased from 0 to 200 nM (lanes 12-18) to study the affect of Gq1 on the PCR amplification of TSR8 which has a sequence identical to the TS primer extended with eight telomeric repeats 5'-(AAT CCG TCG AGC AGA GTT AG (GGT TAG)₇)-3'. Lane 1-3 show the extended telomerase product in the absence of Gq1 with the internal PCR control marked as IC. The concentration of Gq1 in the telomerase extension reaction is varied from 0 nM to 200 nM (Lanes 4-9). Lane 10 is a heat control and lane 11 is a PCR control carried out at 1 μM Gq1.

Figure 16A shows a HeLa cell 48 hours after transfection with Control GFP plasmid pEGFP-N3. Green fluorescence is evenly distributed throughout the cell, in both cytoplasm and nucleus. (B) Schematic diagram to indicate approximate location of cytoplasm (C) and nucleus (N).

Figure 17 shows a single Hela cell transfected with pGq1-NLS-EGFP plasmid, viewed after 48 hours. Fluorescence microscopy (panel A) and phase contrast microscopy (panel B) indicate that the zinc finger-GFP fusion is entirely in the nuclear compartment. Within the nucleus, the GFP is predominantly concentrated within spherical subdomains. Note the multi-lobed nuclear phenotype, indicative of apoptosis.

Figure 18 shows further examples of single HeLa cells (A and B) transfected with pGq1-NLS-EGFP plasmid, viewed by fluorescence microscopy after 48 hours. Again, the zinc finger-GFP fusion is entirely in the nuclear compartment (green). Also, within the nucleus, the GFP is predominantly concentrated within spherical subdomains. Note the multi-lobed nuclear phenotype, indicative of apoptosis.

Figure 19 shows examples of single prometaphase COS7 cells, transfected with pEGFP-Gq1-NLS, and viewed by fluorescence microscopy after 24 hours. Note that the zinc finger localisation is entirely nuclear (green). Propidium iodide staining reveals condensed chromosomes (red).

Figure 20 shows examples of single metaphase (colcemide-treated) COS7 cells, transfected with pEGFP-Gq1-NLS, and viewed by fluorescence microscopy after 24 hours. Note that the zinc finger-EGFP localisation is entirely nuclear (green). Propidium iodide staining reveals condensed chromosomes (red).

Figure 21 shows results of a fluorescence quenching assay to screen for small molecules that bind telomeric DNA sequences. The zinc finger (Gq1) and the telomeric DNA (T) are linked to donor (● : fluorescein) and acceptor (○ : tetramethylrhodamine) molecules for FRET. A: The zinc finger (Gq1) binds to the telomeric DNA (T) such that fluorescence is quenched. The potential drug candidate (D₁) does not displace Gq1 and so no fluorescence is detected. B: The zinc finger (Gq1) is displaced from the telomeric DNA (T) by the potential drug candidate (D₂), such that fluorescence is detected. All reactions are carried out in 384-well plates.

DETAILED DESCRIPTION OF THE INVENTION

Disclosed herein are nucleic acid-binding, preferably DNA binding, polypeptide molecule(s) capable of binding to telomeric G-quadruplex structure(s), and the engineering of these. Preferably, these molecules are polypeptides comprising a zinc finger motif.

Nucleic acid binding polypeptides according to the present invention advantageously bind to single stranded human telomeric DNA with an affinity comparable to the binding of naturally occurring transcription factors to their cognate duplex DNA recognition site(s). DNA in the bound complexes is preferably in the G-quadruplex
5 conformation. The nucleic acid binding polypeptides are capable of binding to their target sequences *in vivo*, and furthermore are capable of inhibiting various enzymatic activities. *In vitro* and *in vivo* assays for enzyme activity are known in the art, and are also set out in the Examples.

As used herein, the term 'isolated or purified' is used to mean that a molecule is
10 free of one or more components of its natural environment. Where the molecule(s) are produced *in vitro* or *in vivo* in a laboratory, they are considered to be isolated or purified. Isolated molecules therefore include such molecules when produced using recombinant cell culture, phage culture etc. Molecules present in an organism expressing a recombinant nucleic acid encoding same, whether the molecule(s) are "isolated" or otherwise, are also
15 included within the scope of the present invention.

The term 'molecule' has its natural meaning. Preferably, such molecules are polypeptides. The expression 'capable of binding to one or more of' is used to indicate that the molecule(s) retain the ability to associate with, interact with, or bind to one or more of the mentioned entities. This binding may be reversible or irreversible. This
20 binding may be temporary or permanent. It may be covalent, ionic, or hydrogen bonding, Van-der-Waals association or any other type of molecular interaction.

Telomeric nucleic acid refers to nucleic acid comprised in or derived from telomeres of eukaryotic cells. The term therefore includes known telomeric repetitive DNA sequences (see below for examples), may include related RNA sequences such as
25 telomeric primer sequences, and may include sub-telomeric repeat sequences, or other sequence(s) found at chromosome ends. The term is intended to include these nucleic acids regardless of their molecular context. This means that such molecules are included if they are in a complex with telomeric or scaffold proteins, or if they are naked *in vitro*. The

molecules are included when they are in vivo such as bona fide telomeres in cell nuclei, or when they are removed from their natural context, such as when on a chef gel or when cloned. The term telomeric nucleic acid may also include mutants, fragments or derivatives thereof, provided such mutants, fragments or derivatives retain substantial
5 sequence homology with said telomeric nucleic acid molecules-this is discussed in more detail below.

Telomeric nucleic acids are known to adopt unconventional or non-conventional structural conformations, mediated by unusual base-pairing (ie. other than simple base paired duplex DNA). Examples of these structures include G-quadruplexes.

10 The term 'G-quadruplex' as understood herein relates to any four-stranded DNA structure. Those skilled in the art realise that these structures comprise loops and hairpins and such like as the two strands of a duplex fold back alongside themselves to form a four-stranded structure, even though only two distinct nucleotide polymer strands may be present. It is also understood that such structures may comprise single-stranded DNA
15 and/or double stranded DNA. Accordingly, in another aspect, the invention relates to a nucleic acid binding molecule as described above wherein said nucleic acid comprises single-stranded DNA. The feature which characterises a 'G-quadruplex' as the term is used herein is that at least a part of the structure to which it refers is in a four-stranded conformation. G-quadruplexes may be intra- or inter- molecular.

20 The term 'G-quartet' refers to that part of a nucleic acid structure which is in a four-stranded conformation. A G-quartet is therefore any segment of nucleic acid or combination of nucleic acids which is in a four-stranded conformation.

Four-stranded nucleic acid conformations (ie. G-quartets) may comprise unconventional base pairing. Conventional base pairing is considered to be Watson and
25 Crick double helical base paired nucleic acid. Unconventional base pairing is therefore base pairing other than Watson and Crick double helical base pairing. Thus, in another

aspect, the invention relates to a nucleic acid binding molecule as described above wherein said nucleic acid is in a non-Watson-Crick base paired conformation.

An example of unconventional base pairing is Hoogsteen base pairing.

The polypeptides preferably comprise a zinc finger motif. A zinc finger is a DNA-
5 binding protein domain that may be used as a scaffold to design DNA-binding proteins. The properties of such motifs include the possession of a Cys2-His2 motif, and are discussed in more detail below. The nucleic acid binding polypeptides provided here preferably exhibit strong discrimination between G-quadruplex nucleic acid and the
10 double-stranded form of the same sequence and between G-quadruplex nucleic acid and the single-stranded variants.

The nucleic acid binding polypeptides described here may be used in screens for molecules which bind to telomeric, G-quadruplex, or G-quartet nucleic acid, or which
15 disrupt the binding between the polypeptides and the nucleic acids. Furthermore, the nucleic acid binding polypeptides may be used in assays for telomerase activity, or telomere length.

This or other aspect(s) may comprise dispensing a nucleic acid sample into the wells of a plate suitable for use with an ELISA reader, such as a 96-well microtitre plate. Gq1* labelled with fluorescent dye or enzyme is then added to the well, incubated and washed, and the binding of the Gq1* molecules to the nucleic acid sample is measured by
20 fluorescence or ELISA. The telomerase or candidate telomerase is added to the nucleic acid sample, and incubated at a suitable temperature for the telomerase or candidate telomerase to function. Fresh Gq1* labelled with fluorescent dye or enzyme is then added to the well, incubated and washed, and the binding of the Gq1* molecules to the nucleic acid sample is measured by fluorescence or ELISA. The binding of the Gq1* molecules to
25 the nucleic acid sample before and after treatment with the telomerase or candidate telomerase is compared. A higher binding coefficient after telomerase treatment indicates that more target nucleic acid is present after telomerase treatment, and thus indicates that

telomerase activity is indeed present in the sample. This method can be easily adapted for estimating the length of telomere(s), by simply measuring the binding of an excess of such nucleic acid binding polypeptides to normalised masses of nucleic acid sample. The amount of bound molecule per given mass of DNA then provides an estimate of the length
5 of the telomere(s), if any are present.

The nucleic acid binding polypeptides are preferably labelled using any suitable method as are well known in the art, including fluorescent labelling, radioactive labelling, peptide tagging, immunolabelling and the like. These are discussed in more detail below.

As described above, the nucleic acid binding polypeptides may be used for
10 discriminating between duplex and quadruplex nucleic acid. 'Discriminating between' means that the two or more entities which are being discriminated may be told apart or mutually excluded or identified or otherwise distinguished. In this example, the term is used to mean that duplex nucleic acid and quadruplex nucleic acid may be distinguished using this method.

15 The nucleic acid binding polypeptides may also be used for manipulating telomeric structure(s) *in vivo*. In this context, 'manipulating' means altering, binding, cleaving, modifying (such as chemical and/or enzymatic modification) or similar effect. An effector domain may be a repressor domain, a nuclease, a tag, an enzyme or enzymatic activity, a toxin, a prodrug or any other suitable effector as discussed below.

20 Exposure of cells to these nucleic acid binding polypeptides results in nuclear localisation. A multilobar-nuclear or multinuclear phenotype is displayed, which we believe is due to apoptotic cell death. Accordingly, we provide the use of nucleic acid binding polypeptides capable of binding to telomeric, G-quadruplex, or G-quartet nucleic acid as cytotoxics, or in methods of killing a cell.

The nucleic acid binding polypeptides preferably have the structures as set out in Figure 3, most preferably, Gq1 of Figure 3. Thus, a highly preferred nucleic acid binding polypeptide for use in the invention comprises the following sequence of binding residues:

Protein	F1	F2	F3
Gq1	DSAHLTR	DRSDLSE	RSDHRIE

5 where the residues shown in the columns F1, F2 and F3 represent residues at positions -1, 1, 2, 3, 4, 5 and 6 respectively of fingers 1, 2 and 3.

A highly preferred embodiment provides a nucleic acid binding polypeptide capable of binding telomeric, G-quadruplex, or G-quartet nucleic acid, which polypeptide (Gq1 polypeptide) comprises the sequence:

10 MAEERPYACPVESCDRRFSDSAHLTRHIRIHTGQKPFQCRICMRNFSDRSDLSEHIRTHT
GEKPFACDICGRKFARSDHRIEHTKIHRLRQKDAAAE

or which polypeptide is encoded by the sequence (Gq1 nucleic acid sequence):

15 ATGGCGGAAGAGAGGCCCTACGCATGCCCTGTCGAGTCCTGCGATCGCCGCTTTTCTGAC
TCGGCCACCTTACCCGGCATATCCGCATCCACACCGGTCAGAAGCCCTTCCAGTGTCTGA
ATCTGCATGCGTAACTTCAGTGACAGGTCCGACCTGAGCGAACACATCCGCACCCACACA
GGCGAGAAGCCTTTTGCTGTGACATTTGTGGGAGGAAATTTGCCCGCAGCGACCACCGC
ATAGAACATACCAAGATACACCTGCGCCAAAAAGATGCGGCCGCGGAG

20 Preferably, the nucleic acid binding polypeptide comprises one finger, any combination of two fingers as set out above, or three fingers as set out above.

In a preferred embodiment of the invention, we provide nucleic acid binding polypeptides which comprise six zinc finger motifs, in which a flexible or structured linker links fingers 3 and 4. Preferably such dimers comprise F1 to F3 of Gq1 joined to F1 to F3 of Gq1. A further embodiment provides a four finger polypeptide with a structured or
25 flexible linker between fingers 2 and 3. An example of this is a polypeptide comprising F1

and F2 of Gq1 linked to F1 and F2 of Gq1. Other combinations of fingers, for example, comprising combinations of fingers F1, F2, and/or F3 are also possible; such polypeptides may comprise 4, 5, 6, 7, 8, 9, 10, 11, 12 or more fingers. The dimeric or polymeric polypeptides may be constructed by linking a two or three finger polypeptide with one or
 5 more other two or three finger polypeptides with a structured or flexible linker. The linkers joining the fingers may comprise canonical or preferably structured or flexible linkers.

Specific constructs and polypeptides according to this aspect of the invention are set out in the Examples. In particular, the invention encompasses the polypeptides, and the use of these, according to the following: Construct Gq1(1:3)-linkerA-Gq1(1:3) comprising
 10 [Gq1 Fingers1-3]- linkerA - [Gq1 Fingers1-3], Construct Gq1(1:3)-linkerB-Gq1(1:3) comprising [Gq1 Fingers1-3] - linkerB - [Gq1 Fingers1-3], Construct Gq1(1:2)-linkerA-Gq1(1:2) comprising [Gq1 Fingers1-2] - linkerA - [Gq1 Fingers1-2], or Construct Gq1(1:2)-linkerB-Gq1(1:2) comprising [Gq1 Fingers1-2] - linkerB - [Gq1 Fingers1-2], where: linkerA = TG GGGG ERP and linkerB = TG GGGG GGS GGS GGS GGS ERP.
 15 The invention also includes the above polypeptides in which the linker comprises TG GGGG GGGG GGGG GGGG GGGG ERP.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA and immunology, which are within the capabilities of a person of ordinary skill in
 20 the art. Such techniques are explained in the literature. See, for example, J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, *Molecular Cloning: A Laboratory Manual*, Second Edition, Books 1-3, Cold Spring Harbor Laboratory Press; Ausubel, F. M. et al. (1995 and periodic supplements; *Current Protocols in Molecular Biology*, ch. 9, 13, and 16, John Wiley & Sons, New York, N.Y.); B. Roe, J. Crabtree, and A. Kahn, 1996, *DNA Isolation and Sequencing: Essential Techniques*, John Wiley & Sons; J. M. Polak and James O'D. McGee, 1990, *In Situ Hybridization: Principles and Practice*; Oxford University Press; M. J. Gait (Editor), 1984, *Oligonucleotide Synthesis: A Practical Approach*, Irl Press; and, D. M. J. Lilley and J. E. Dahlberg, 1992, *Methods of Enzymology: DNA Structure Part A: Synthesis and Physical Analysis of DNA* Methods in Enzymology, Academic Press. Each
 25 of these general texts is herein incorporated by reference.
 30

NUCLEIC ACID BINDING POLYPEPTIDES

The present invention relates in one aspect to the production and use of nucleic acid binding polypeptides. Such nucleic acid binding polypeptides are preferably engineered. The term "engineered" means that the nucleic acid binding polypeptide, zinc finger peptide, polypeptide, protein or fusion protein has been generated or modified *in vitro*. Typically a zinc finger polypeptide is produced by deliberate mutagenesis, for example the substitution of one or more amino acid residues, either as part of a random mutagenesis procedure or by site-directed mutagenesis, or by selection from a library or libraries of mutated zinc finger peptides. Engineered zinc finger peptides for use in the invention can also be produced *de novo* using rational design strategies.

The term "polypeptide", "peptide" and "protein" are used interchangeably to refer to a polymer of amino acid residues, preferably including naturally occurring amino acid residues. Artificial analogues of amino acids may also be used in the nucleic acid binding polypeptides, to impart the proteins with desired properties or for other reasons. Thus, the term "amino acid", particularly in the context where "any amino acid" is referred to, means any sort of natural or artificial amino acid or amino acid analogue that may be employed in protein construction according to methods known in the art. Moreover, any specific amino acid referred to herein may be replaced by a functional analogue thereof, particularly an artificial functional analogue. Polypeptides may be modified, for example by the addition of carbohydrate residues to form glycoproteins. The nomenclature used herein therefore specifically comprises within its scope functional analogues or mimetics of the defined amino acids.

As used herein, "nucleic acid" includes both RNA and DNA, constructed from natural nucleic acid bases or synthetic bases, or mixtures thereof. Preferably, however, the binding polypeptides of the invention are DNA binding polypeptides.

Zinc Fingers

Particularly preferred examples of nucleic acid binding polypeptides are zinc finger peptides. Zinc finger peptides typically contain strings of small domains, known as "fingers", each stabilised by the co-ordination of zinc. Thus, binding of zinc finger polypeptides to target nucleic acid sequences occurs via α -helical zinc metal atom co-ordinated binding motifs known as zinc fingers. Zinc fingers are capable of recognising and binding to a nucleic acid triplet, or an overlapping quadruplet, in a nucleic acid binding sequence. Particularly preferred nucleic acid binding polypeptides comprise zinc fingers of the Cys2-His2 type.

10 However, zinc fingers are also known to bind RNA and proteins (Searles, M. A. *et al.*, *J. Mol. Biol.* 301: 47-60 (2000); Mackay, J. P. & Crossley, M. *Trends Biochem. Sci.* 23: 1-4).

Preferably, there are 2 or more zinc fingers, for example 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 or more zinc fingers, in each zinc finger polypeptide.

15 Advantageously, the zinc finger polypeptide comprises 3 or more zinc fingers. Furthermore, the number of zinc fingers in a zinc finger polypeptide is preferably a multiple of two.

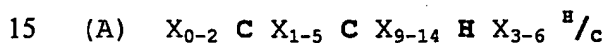
The DNA binding residue positions of zinc finger polypeptides, as referred to herein, are numbered from the first residue in the α -helix of the finger, ranging from +1 to +9. "-1" refers to the residue in the framework structure immediately preceding the α -helix in a Cys2-His2 zinc finger polypeptide. Residues referred to as "++" are residues present in an adjacent (C-terminal) finger. Where there is no C-terminal adjacent finger, "++" interactions do not operate.

25 The α -helix of a zinc finger binding protein aligns antiparallel to the nucleic acid strand, such that the primary nucleic acid sequence is arranged 3' to 5' in order to correspond with the N- terminal to C-terminal sequence of the zinc finger. Since nucleic acid sequences are conventionally written 5' to 3', and amino acid sequences N-terminus

to C-terminus, the result is that when a nucleic acid sequence and a zinc finger polypeptide are aligned according to convention, the primary interaction of the zinc finger is with the - strand of the nucleic acid, since it is this strand which is aligned 3' to 5'. These conventions are followed in the nomenclature used herein. It should be noted, however, that in nature certain fingers, such as finger 4 of the protein GLI, bind to the + strand of the nucleic acid sequence. See Suzuki *et al.* (1994) *Nucl. Acids Rev.* 22: 3397-3405; and Pavletich and Pabo, (1993) *Science* 261: 1701-1707. The present invention encompasses incorporation of such zinc finger peptides into DNA binding molecules.

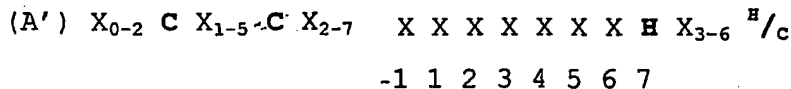
A zinc finger binding motif is a structure well known to those in the art and defined in, for example, Miller *et al.*, (1985) *EMBO J.* 4:1609-1614; Berg (1988) *PNAS (USA)* 85:99-102; Lee *et al.*, (1989) *Science* 245:635-637; see International patent applications WO 96/06166 and WO 96/32475, corresponding to USSN 08/422,107, incorporated herein by reference.

In general, a preferred zinc finger framework has the structure:



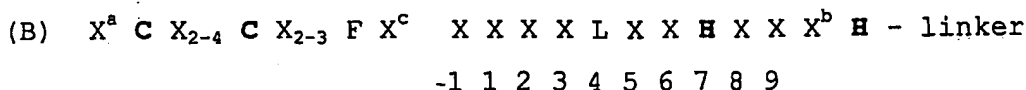
where X is any amino acid, and the numbers in subscript indicate the possible numbers of residues represented by X.

The above framework may be further refined to include the structure:



where X is any amino acid, and the numbers in subscript indicate the possible numbers of residues represented by X.

In a preferred aspect, zinc finger nucleic acid binding motifs may be represented as motifs having the following primary structure:



wherein X (including X^a , X^b and X^c) is any amino acid. X_{2-4} and X_{2-3} refer to the presence of 2 or 4, or 2 or 3, amino acids, respectively.

The Cys and His residues, which together co-ordinate the zinc metal atom, are marked in bold text and are usually invariant, as is the Leu residue at position +4 in the α -helix. Residues X, X^a , X^b , X^c etc are referred to for convenience as "backbone" residues.

Modifications to the standard representation of a zinc finger may occur or be effected without necessarily abolishing zinc finger peptide function, by insertion, mutation or deletion of amino acid residues. For example the second His residue may be replaced by Cys (Krizek *et al.* (1991) *J. Am. Chem. Soc.* 113: 4518-4523) and that Leu at +4 can in some circumstances be replaced with Arg. The Phe residue before X_c may be replaced by any aromatic residue other than Trp. Moreover, experiments have shown that departure from the preferred structure and residue assignments for a zinc finger peptide are tolerated and may even prove beneficial in binding to certain nucleic acid sequences. Even taking this into account, however, the general structure involving an α -helix co-ordinated by a zinc atom which contacts four Cys or His residues, is not altered. As used herein, structures (A), (A') and (B) above are taken as an exemplary structure representing all zinc finger peptide structures of the Cys2-His2 type.

Preferably, X^a is $F/Y-X$ or $P-F/Y-X$. In this context, X is any amino acid. Preferably, in this context X is E, K, T or S. Less preferred but also envisaged are Q, V, A and P. The remaining amino acids remain possible.

Preferably, X_{2-4} consists of two amino acids rather than four. The first of these amino acids may be any amino acid, but S, E, K, T, P and R are preferred. Advantageously, it is P or R. The second of these amino acids is preferably E, although any amino acid may be used.

Preferably, X^b is T or I. Preferably, X^c is S or T.

Preferably, $X_{2,3}$ is G-K-A, G-K-C, G-K-S or G-K-G. However, departures from the preferred residues are possible, for example in the form of M-R-N or M-R.

The linker may comprise a sequence T-G-^E/Q-^K/R or T-G-^E/Q-^K/R-P. The linker
5 may comprise a canonical, structured or flexible linker. Structured and flexible linkers (as well as canonical linkers) are described elsewhere in this document, and in our UK application numbers GB 0001582.6, GB0013103.7, GB0013104.5 and our International Patent Application PCT/GB00/00202, all of which are hereby incorporated by reference.

Engineering, Rational and Rule Based Design of Zinc Fingers

10 The rules set forth for zinc finger polypeptide design in our European or PCT patent applications having publication numbers WO 98/53057, WO 98/53060, WO 98/53058, WO 98/53059 may be used to design zinc finger proteins for use in the present invention. These publications describe improved techniques for designing zinc finger polypeptides capable of binding desired nucleic acid sequences. Engineering of zinc
15 fingers which involves applying rules which specify the choice of amino acid residues based on the identity of residues in a target nucleic acid sequence is referred to here as "rule based" or "rational" design. Such rational design provides a great deal of versatility in zinc finger design.

In combination with selection procedures, such as phage display, set forth for
20 example in WO 96/06166 and described in further detail below, these techniques enable the production of zinc finger polypeptides capable of recognising practically any desired sequence.

The zinc finger polypeptides for use in the present invention may be produced using a method for preparing a nucleic acid binding protein of the Cys2-His2 zinc finger
25 class capable of binding to a nucleic acid triplet in a target nucleic acid sequence, wherein binding to each base of the triplet by an α -helical zinc finger nucleic acid binding motif in

the protein is determined as follows: (a) if the 5' base in the triplet is G, then position +6 in the α -helix is Arg; or position +6 is Ser or Thr and position ++2 is Asp; (b) if the 5' base in the triplet is A, then position +6 in the α -helix is Gln and ++2 is not Asp; (c) if the 5' base in the triplet is T, then position +6 in the α -helix is Ser or Thr and position ++2 is Asp; (d) if the 5' base in the triplet is C, then position +6 in the α -helix may be any amino acid, provided that position ++2 in the α -helix is not Asp; (e) if the central base in the triplet is G, then position +3 in the α -helix is His; (f) if the central base in the triplet is A, then position +3 in the α -helix is Asn; (g) if the central base in the triplet is T, then position +3 in the α -helix is Ala, Ser or Val; provided that if it is Ala, then one of the residues at -1 or +6 is a small residue; (h) if the central base in the triplet is C, then position +3 in the α -helix is Ser, Asp, Glu, Leu, Thr or Val; (i) if the 3' base in the triplet is G, then position -1 in the α -helix is Arg; (j) if the 3' base in the triplet is A, then position -1 in the α -helix is Gln; (k) if the 3' base in the triplet is T, then position -1 in the α -helix is Asn or Gln; (l) if the 3' base in the triplet is C, then position -1 in the α -helix is Asp.

Furthermore, a nucleic acid binding protein of the Cys2-His2 zinc finger class capable of binding to a nucleic acid quadruplet in a target nucleic acid sequence comprising a target nucleotide sequence may be prepared using the following rules. Binding to each base of the quadruplet by an α -helical zinc finger nucleic acid binding motif in the protein is determined as follows: (a) if base 4 in the quadruplet is G, then position +6 in the α -helix is Arg or Lys; (b) if base 4 in the quadruplet is A, then position +6 in the α -helix is Glu, Asn or Val; (c) if base 4 in the quadruplet is T, then position +6 in the α -helix is Ser, Thr, Val or Lys; (d) if base 4 in the quadruplet is C, then position +6 in the α -helix is Ser, Thr, Val, Ala, Glu or Asn; (e) if base 3 in the quadruplet is G, then position +3 in the α -helix is His; (f) if base 3 in the quadruplet is A, then position +3 in the α -helix is Asn; (g) if base 3 in the quadruplet is T, then position +3 in the α -helix is Ala, Ser or Val; provided that if it is Ala, then one of the residues at -1 or +6 is a small residue; (h) if base 3 in the quadruplet is C, then position +3 in the α -helix is Ser, Asp, Glu, Leu, Thr or Val; (i) if base 2 in the quadruplet is G, then position -1 in the α -helix is Arg; (j) if

base 2 in the quadruplet is A, then position -1 in the α -helix is Gln; (k) if base 2 in the quadruplet is T, then position -1 in the α -helix is His or Thr; (l) if base 2 in the quadruplet is C, then position -1 in the α -helix is Asp or His; (m) if base 1 in the quadruplet is G, then position +2 is Glu; (n) if base 1 in the quadruplet is A, then position +2 Arg or Gln; (o) if
 5 base 1 in the quadruplet is C, then position +2 is Asn, Gln, Arg, His or Lys; (p) if base 1 in the quadruplet is T, then position +2 is Ser or Thr.

The above rules may be further refined, to provide a method for preparing a nucleic acid binding protein of the Cys2-His2 zinc finger class capable of binding to a nucleic acid quadruplet in a target nucleic acid sequence comprising a target nucleotide
 10 sequence, wherein binding to each base of the quadruplet by an α -helical zinc finger nucleic acid binding motif in the protein is determined as follows: (a) if base 4 in the quadruplet is G, then position +6 in the α -helix is Arg; or position +6 is Ser or Thr and position ++2 is Asp; (b) if base 4 in the quadruplet is A, then position +6 in the α -helix is Gln and ++2 is not Asp; (c) if base 4 in the quadruplet is T, then position +6 in the α -helix
 15 is Ser or Thr and position ++2 is Asp; (d) if base 4 in the quadruplet is C, then position +6 in the α -helix may be any amino acid, provided that position ++2 in the α -helix is not Asp; (e) if base 3 in the quadruplet is G, then position +3 in the α -helix is His; (f) if base 3 in the quadruplet is A, then position +3 in the α -helix is Asn; (g) if base 3 in the quadruplet is T, then position +3 in the α -helix is Ala, Ser or Val; provided that if it is Ala,
 20 then one of the residues at -1 or +6 is a small residue; (h) if base 3 in the quadruplet is C, then position +3 in the α -helix is Ser, Asp, Glu, Leu, Thr or Val; (i) if base 2 in the quadruplet is G, then position -1 in the α -helix is Arg; (j) if base 2 in the quadruplet is A, then position -1 in the α -helix is Gln; (k) if base 2 in the quadruplet is T, then position -1 in the α -helix is Asn or Gln; (l) if base 2 in the quadruplet is C, then position -1 in the α -
 25 helix is Asp; (m) if base 1 in the quadruplet is G, then position +2 is Asp; (n) if base 1 in the quadruplet is A, then position +2 is not Asp; (o) if base 1 in the quadruplet is C, then position +2 is not Asp; (p) if base 1 in the quadruplet is T, then position +2 is Ser or Thr.

As set out above, the major binding interactions occur with amino acids -1, +3 and +6. Amino acids +4 and +7 are largely invariant. The remaining amino acids may be

essentially any amino acids. Preferably, position +9 is occupied by Arg or Lys. Advantageously, positions +1, +5 and +8 are not hydrophobic amino acids, that is to say are not Phe, Trp or Tyr. Preferably, position ++2 is any amino acid, and preferably serine, save where its nature is dictated by its role as a ++2 amino acid for an N-terminal zinc finger in the same nucleic acid binding molecule.

The foregoing represents sets of rules which permits the design of a zinc finger binding protein specific for any given target DNA sequence. In a most preferred aspect, therefore, the above rules allow the definition of every residue in a zinc finger peptide DNA binding motif which will bind specifically to a given target DNA triplet or quadruplet. In order to produce a binding protein having improved binding, moreover, the rules described here may be supplemented by physical or virtual modelling of the protein/DNA interface in order to assist in residue selection.

The code provided by the present invention is not entirely rigid; certain choices are provided. For example, positions +1, +5 and +8 may have any amino acid allocation, whilst other positions may have certain options: for example, the present rules provide that, for binding to a central T residue, any one of Ala, Ser or Val may be used at +3. In its broadest sense, therefore, the present invention provides a very large number of proteins which are capable of binding to every defined target DNA triplet.

Preferably, however, the number of possibilities may be significantly reduced. For example, the non-critical residues +1, +5 and +8 may be occupied by the residues Lys, Thr and Gln respectively as a default option. In the case of the other choices, for example, the first-given option may be employed as a default. Thus, the code according to the present invention allows the design of a single, defined polypeptide (a "default" polypeptide) which will bind to its target triplet. Zinc fingers may be based on naturally occurring zinc fingers and consensus zinc fingers.

Accordingly, the zinc finger polypeptides described and for use here can be prepared using a method comprising the steps of: (a) selecting a model zinc finger peptide

from the group consisting of naturally occurring zinc finger proteins and consensus zinc finger peptides; and (b) mutating at least one of positions -1, +3, +6 (and ++2) of the peptide.

In general, naturally occurring zinc fingers may be selected from those fingers for which the DNA binding specificity is known. For example, these may be the fingers for which a crystal structure has been resolved: namely Zif 268 (Elrod-Erickson *et al.*, (1996) Structure 4:1171-1180), GLI (Pavletich and Pabo, (1993) Science 261:1701-1707), Tramtrack (Fairall *et al.*, (1993) Nature 366:483-487) and YY1 (Houbaviy *et al.*, (1996) PNAS (USA) 93:13577-13582). Preferably, the modified nucleic acid binding polypeptide is derived from Zif 268, GAC, or a Zif-GAC fusion comprising three fingers from Zif linked to three fingers from GAC. By "GAC-clone", we mean a three-finger variant of ZIF268 which is capable of binding the sequence GCGGACGCG, as described in Choo & Klug (1994), *Proc. Natl. Acad. Sci. USA*, 91, 11163-11167.

Although mutation of the DNA-contacting amino acid residues of the DNA binding domain of zinc finger peptides allows selection of peptides which bind to desired target nucleic acids, in a preferred embodiment residues which are outside the DNA-contacting region may be mutated. Mutations in such residues may affect the interaction between zinc finger peptides in a zinc finger polypeptide, and thus alter binding site specificity. For instance, Arg at the +10 position of TFIIIA finger 3 makes a base specific contact to guanine (Nolte, R. T. *et al.*, *Proc. Natl. Acad. Sci. USA* 95: 2938-2943 (1998). Similarly, residues other than those at positions -1, +3, +6 and ++2 may also be utilised for binding RNA molecules.

The naturally occurring zinc finger 2 in Zif 268 makes an excellent starting point from which to engineer a zinc finger and is preferred.

Consensus zinc finger structures may be prepared by comparing the sequences of known zinc fingers, irrespective of whether their binding domain is known. Preferably, the consensus structure is selected from the group consisting of the consensus structure P Y K

C P E C G K S F S Q K S D L V K H Q R T H T , and the consensus structure P Y K C S
E C G K A F S Q K S N L T R H Q R I H T. The consensuses are derived from the
consensus provided by Krizek *et al.*, (1991) J. Am. Chem. Soc. 113: 4518-4523 and from
Jacobs, (1993) PhD thesis, University of Cambridge, UK. In both cases, canonical,
5 structured or flexible linker sequences, as described below, may be formed on the ends of
the consensus for joining two zinc finger domains together.

When the nucleic acid specificity of the model finger selected is known, the
mutation of the finger in order to modify its specificity to bind to the target DNA may be
directed to residues known to affect binding to bases at which the natural and desired
10 targets differ. Otherwise, mutation of the model fingers should be concentrated upon
residues -1, +3, +6 and ++2 as provided for in the foregoing rules.

Selection of Zinc Fingers from Libraries

The rational design described above may be used instead of, or to complement zinc
finger production by selection from libraries.

15 Thus, the zinc finger polypeptides described here and capable of binding to a target
DNA sequence comprising a target nucleotide sequence may be produced by a method
comprising: a) providing a nucleic acid library encoding a repertoire of zinc finger
domains or modules, the nucleic acid members of the library being at least partially
randomised at one or more of the positions encoding residues -1, 2, 3 and 6 of the α -helix
20 of the zinc finger modules; b) displaying the library in a selection system and screening it
against the target DNA sequence; and c) isolating the nucleic acid members of the library
encoding zinc finger modules or domains capable of binding to the target sequence.

The term "library" is used according to its common usage in the art, to denote a
collection of polypeptides or, preferably, nucleic acids encoding polypeptides. These
25 polypeptides contain regions of randomisation, such that each library will comprise or
encode a repertoire of polypeptides, wherein individual polypeptides differ in sequence
from each other. The same principle is present in virtually all libraries developed for

selection, such as by phage display. Methods for the production of libraries encoding randomised members such as polypeptides are known in the art and may be applied in the present invention. The members of the library may contain regions of randomisation, such that each library will comprise or encode a repertoire of polypeptides, wherein individual polypeptides differ in sequence from each other. The same principle is present in virtually all libraries developed for selection, such as by phage display.

Randomisation, as used herein, refers to the variation of the sequence of the polypeptides which comprise the library, such that various amino acids may be present at any given position in different polypeptides. Randomisation may be complete, such that any amino acid may be present at a given position, or partial, such that only certain amino acids are present. Preferably, the randomisation is achieved by mutagenesis at the nucleic acid level, for example by synthesising novel genes encoding mutant proteins and expressing these to obtain a variety of different proteins. Alternatively, existing genes can be themselves mutated, such by site-directed or random mutagenesis, in order to obtain the desired mutant genes.

Zinc finger polypeptides may be designed which specifically bind to nucleic acids incorporating the base U, in preference to the equivalent base T.

A further method for producing a zinc finger polypeptide for use here and capable of binding to a target DNA sequence comprising a telomeric, G-quadruplex, or G-quartet nucleic acid comprises: a) providing a nucleic acid library encoding a repertoire of zinc finger polypeptides each possessing more than one zinc finger, the nucleic acid members of the library being at least partially randomised at one or more of the positions encoding residues -1, 2, 3 and 6 of the α -helix in a first zinc finger and at one or more of the positions encoding residues -1, 2, 3 and 6 of the α -helix in a further zinc finger of the zinc finger polypeptides; b) displaying the library in a selection system and screening it against the target DNA sequence; and d) isolating the nucleic acid members of the library encoding zinc finger polypeptides capable of binding to the target sequence.

In this aspect, the invention encompasses library technology described in our International patent application WO 98/53057, incorporated herein by reference in its entirety. WO 98/53057 describes the production of zinc finger polypeptide libraries in which each individual zinc finger polypeptide comprises more than one, for example two or three, zinc fingers; and wherein within each polypeptide partial randomisation occurs in at least two zinc fingers. This allows for the selection of the "overlap" specificity, wherein, within each triplet, the choice of residue for binding to the third nucleotide (read 3' to 5' on the + strand) is influenced by the residue present at position +2 on the subsequent zinc finger, which displays cross-strand specificity in binding. The selection of zinc finger polypeptides incorporating cross-strand specificity of adjacent zinc fingers enables the selection of nucleic acid binding proteins more quickly, and/or with a higher degree of specificity than is otherwise possible.

Thus, zinc finger binding motifs designed accordingly may be combined into nucleic acid binding polypeptide molecules having a multiplicity of zinc fingers. Preferably, the proteins have at least two zinc fingers. The presence of at least three zinc fingers is preferred. Nucleic acid binding proteins may be constructed by joining the required fingers end to end, N-terminus to C-terminus, with canonical, flexible or structured linkers, as described elsewhere. Preferably, this is effected by joining together the relevant nucleic acid sequences which encode the zinc fingers to produce a composite nucleic acid coding sequence encoding the entire binding protein. A "leader" peptide may be added to the N-terminal finger. Preferably, the leader peptide comprises MAEEKP, MAEKP, MAERP or MAEERP. Other polypeptide motifs may be added as desired, for example, nuclear localisation sequences, transcriptional modulator domains such as repressor domains or activation domains, etc.

We therefore describe a method for producing a DNA binding protein for use as described here, wherein the DNA binding protein is constructed by recombinant DNA technology, the method comprising the steps of: preparing a nucleic acid coding sequence encoding a plurality of zinc finger domains or modules defined above, inserting the nucleic acid sequence into a suitable expression vector; and expressing the nucleic acid sequence in a host organism in order to obtain the DNA binding protein.

Flexible and Structured Linkers

The nucleic acid binding polypeptides described here may comprise one or more linker sequences. The linker sequences may comprise one or more flexible linkers, one or more structured linkers, or any combination of flexible and structured linkers. Such linkers
5 are disclosed in our co-pending British Patent Application Numbers 0001582.6, 0013102.9, 0013103.7, 0013104.5 and International Patent Application Number PCT/GB01/00202, which are incorporated by reference.

By "linker sequence" we mean an amino acid sequence that links together two nucleic acid binding modules. For example, in a "wild type" zinc finger protein, the linker
10 sequence is the amino acid sequence lacking secondary structure which lies between the last residue of the α -helix in a zinc finger and the first residue of the β -sheet in the next zinc finger. The linker sequence therefore joins together two zinc fingers. Typically, the last amino acid in a zinc finger is a threonine residue, which caps the α -helix of the zinc finger, while a tyrosine/phenylalanine or another hydrophobic residue is the first amino
15 acid of the following zinc finger. Accordingly, in a "wild type" zinc finger, glycine is the first residue in the linker, and proline is the last residue of the linker. Thus, for example, in the Zif268 construct, the linker sequence is G(E/Q)(K/R)P.

A "flexible" linker is an amino acid sequence which does not have a fixed structure (secondary or tertiary structure) in solution. Such a flexible linker is therefore free to adopt
20 a variety of conformations. An example of a flexible linker is the canonical linker sequence GERP/GEKP/GQRP/GQKP. Flexible linkers are also disclosed in WO99/45132 (Kim and Pabo). By "structured linker" we mean an amino acid sequence which adopts a relatively well-defined conformation when in solution. Structured linkers are therefore those which have a particular secondary and/or tertiary structure in solution.

25 Determination of whether a particular sequence adopts a structure may be done in various ways, for example, by sequence analysis to identify residues likely to participate in protein folding, by comparison to amino acid sequences which are known to adopt certain conformations (e.g., known α -helix, β -sheet or zinc finger sequences), by NMR

spectroscopy, by X-ray diffraction of crystallised peptide containing the sequence, etc as known in the art.

The structured linkers of our invention preferably do not bind nucleic acid, but where they do, then such binding is not sequence specific. Binding specificity may be
5 assayed for example by gel-shift as described below.

The linker may comprise any amino acid sequence that does not substantially hinder interaction of the nucleic acid binding modules with their respective target subsites. Preferred amino acid residues for flexible linker sequences include, but are not limited to, glycine, alanine, serine, threonine proline, lysine, arginine, glutamine and glutamic acid..

10 The linker sequences between the nucleic acid binding domains preferably comprise five or more amino acid residues. The flexible linker sequences according to our invention consist of 5 or more residues, preferably, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 or more residues. In a highly preferred embodiment of the invention, the flexible linker sequences consist of 5, 7 or 10 residues.

15 Once the length of the amino acid sequence has been selected, the sequence of the linker may be selected, for example by phage display technology (see for example United States Patent No. 5,260,203) or using naturally occurring or synthetic linker sequences as a scaffold (for example, GQKP and GEKP, see Liu et al., 1997, *Proc. Natl. Acad. Sci. USA* 94, 5525-5530 and Whitlow et al., 1991, *Methods: A Companion to Methods in*
20 *Enzymology* 2: 97-105). The linker sequence may be provided by insertion of one or more amino acid residues into an existing linker sequence of the nucleic acid binding polypeptide. The inserted residues may include glycine and/or serine residues. Preferably, the existing linker sequence is a canonical linker sequence selected from GEKP, GERP, GQKP and GQRP. More preferably, each of the linker sequences comprises a sequence
25 selected from GGEKP, GGQKP, GGSGEKP, GGSGQKP, GGSGGSGEKP, and GGSGGSQKP.

Structured linker sequences are typically of a size sufficient to confer secondary or tertiary structure to the linker; such linkers may be up to 30, 40 or 50 amino acids long. In a preferred embodiment, the structured linkers are derived from known zinc fingers which do not bind nucleic acid, or are not capable of binding nucleic acid specifically. An
5 example of a structured linker of the first type is TFIIIA finger IV; the crystal structure of TFIIIA has been solved, and this shows that finger IV does not contact the nucleic acid (Nolte *et al.*, 1998, *Proc. Natl. Acad. Sci. USA* 95, 2938-2943.). An example of the latter type of structured linker is a zinc finger which has been mutagenised at one or more of its base contacting residues to abolish its specific nucleic acid binding capability. Thus, for
10 example, a ZIF finger 2 which has residues -1, 2, 3 and 6 of the recognition helix mutated to serines so that it no longer specifically binds DNA may be used as a structured linker to link two nucleic acid binding domains.

The use of structured or rigid linkers to jump the minor groove of DNA is likely to be especially beneficial in (i) linking zinc fingers that bind to widely separated (>3bp)
15 DNA sequences, and (ii) also in minimising the loss of binding energy due to entropic factors.

Typically, the linkers are made using recombinant nucleic acids encoding the linker and the nucleic acid binding modules, which are fused via the linker amino acid sequence. The linkers may also be made using peptide synthesis and then linked to the
20 nucleic acid binding modules. Methods of manipulating nucleic acids and peptide synthesis methods are known in the art (see, for example, Maniatis, *et al.*, 1991. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press).

Zinc finger peptides may also be linked non-covalently. Non-covalent dimerisation
25 domains such as leucine zippers, and coiled coils are preferable for this purpose (O'Shea, *Science*, 254: 539 (1991); Klemm *et al.*, *Ann. Rev. Immunol.* 16: 569-592 (1998); Ho, *et al.*, *Nature*, 382: 822-826 (1996); Pomeranz, *et al.*, *Biochem.* 37: 965 (1998).

Chimeric Nucleic Acid Binding Polypeptides

A chimeric nucleic acid binding polypeptide comprises a nucleotide binding domain (comprising a number of nucleic acid binding polypeptide modules or fingers) designed to bind specifically to a nucleotide sequence, together with one or more further biological effector domains. The term "biological effector domain" should be taken to mean any polypeptide that has a biological function. Included are enzymes, receptors, regulatory domains, activation or repression domains, binding sequences, dimerisation, trimerisation or multimerisation sequences, sequences involved in protein transport, localisation sequences such as subcellular localisation sequences, nuclear localisation, protein targeting or signal sequences. Furthermore, biological effector domains may comprise polypeptides involved in chromatin remodelling, chromatin condensation or decondensation, DNA replication, transcription, translation, protein synthesis, etc. Fragments of such polypeptides comprising the relevant activity are also included in this definition.

The effector domain(s) may be covalently or non-covalently attached to the nucleotide-binding domain.

Chimeric nucleic acid binding polypeptides preferably comprise transcription factor activity, for example, a transcriptional modulation activity such as transcriptional activator or transcriptional repressor activity. For example, a zinc finger chimera protein may comprise a nucleotide binding domain designed to bind specifically to a particular nucleotide sequence, and one or more further biological effector domains, preferably a transcriptional activator or repressor domain, as described in further detail below. The zinc finger chimera may comprise one or more zinc fingers or zinc finger binding modules.

Preferably, in the case of a chimera comprising transcriptional modulation activity, a nuclear localization domain is attached to the DNA binding domain to direct the chimera to the nucleus.

Generally, the nucleic acid binding polypeptide chimera such as a zinc finger chimera peptide may also include an effector domain to regulate gene expression. The effector domain may be directly derived from a basal or regulated transcription factor such as a transactivator, repressor, insulator or silencer (Choo & Klug (1995) *Curr. Opin. Biotech.* 6: 431-436; Choo & Klug (1997); Rebar & Pabo (1994) *Science* 263: 671-673; Jamieson *et al.* (1994) *Biochem.* 33: 5689-5695; Goodrich *et al.*, *Cell* 84: 825-830 (1996); CTCF (Vostrov, A. A. & Quitschke, W. W. *J. Biol. Chem.* 272: 33353-33359 (1997)). Other useful domains may be derived from membrane receptors such as nuclear hormone receptors (Kumar, R & Thompson, E. B. *Steroids* 64: 310-319 (1999)), and their co-activators and co-repressors (Ugai, H. *et al.*, *J. Mol. Med.* 77: 481-494 (1999)).

The nucleic acid binding polypeptide chimera such as a zinc finger chimera protein may also preferably include other domains that may be advantageous within the context of the control of gene expression. These domains may include protein-modifying domains such as histone acetyltransferases, kinases and phosphatases, which can silence or activate genes by modifying DNA structure or the proteins that associate with nucleic acids (Wolffe, *Science* 272: 371-372 (1996); Taunton *et al.*, *Science* 272: 408-411 (1996); Hassig *et al.*, *Proc. Natl. Acad. Sci. USA* 95: 3519-3524 (1998); Wang, *Trends Biochem. Sci.* 19: 373-376 (1994); and Schonthal & Semin, *Cancer Biol.* 6: 239-248 (1995)). Additional effector domains, which can be useful in the context of the present invention are those that modify or rearrange nucleic acid molecules such as methyltransferases, endonucleases, ligases, recombinases etc. (Wood, *Ann. Rev. Biochem.* 65: 135-167 (1996); Sadowski, *FASEB J.* 7: 760-767 (1993); Cheng, *Curr. Opin. Struct. Biol.* 5: 4-10 (1995)) (Wu *et al.* (1995) *Proc. Natl. Acad. Sci. USA* 92:344-348; Nahon & Raveh (1998); Smith *et al.* (1999); and Carroll *et al.* (1999)). It will be appreciated that the biological effector domain portion of the chimera may itself also comprise such activities, without the need for further domains.

In one embodiment, the VP64 domain from herpes simplex virus (HSV) is used to activate gene expression (Seipel *et al.*, *EMBO J.* 11: 4961-4968 (1996)). Other preferred transactivator domains include the HSV VP16 domain (Hagmann *et al.*, *J. Virol.* 71: 5952-5962 (1997)), transactivation domain 1 and / or domain 2 of the p65 subunit of nuclear

factor- κ B (NF- κ B, Schmitz, M. L. *et al.*, *J. Biol. Chem.* 270: 15576-15584 (1995)). Other transcription factors are reviewed in, for example, Lekstrom-Himes J. & Xanthopoulos K. G. (C/EBP family, *J. Biol. Chem.* 273: 28545-28548 (1998)), Bieker, J. J. *et al.*, (globin gene transcription factors, *Ann. N. Y. Acad. Sci.* 850: 64-69 (1998), and Parker, M. G. (oestrogen receptors, *Biochem. Soc. Symp.* 63: 45-50 (1998)).

Use of a transactivation domain from the estrogen receptor is disclosed in Metivier, R., Petit, FG., Valotaire, Y. & Pakdel, F. (2000) *Mol. Endocrinol.* 14: 1849-1871. Furthermore, activation domains from the globin transcription factors EKLF (Pandya, K. Donze, D. & Townes T. (2001) *J. Biol. Chem.* 276: 8239-8243) may also be used, as well as a transactivation domain from FKLF (Asano, H. Li, XS. & Stamatoyannopoulos, G. (1999) *Mol. Cell. Biol.* 19: 3571-3579). C/EPB transactivation domains may also be employed in the methods described here. The C/EBP epsilon activation domain is disclosed in Verbeek, W., Gombart, AF, Chumakov, AM, Muller, C, Friedman, AD, & Koeffler, HP (1999) *Blood* 15: 3327-3337. Kowenz-Leutz, E. & Leutz, A. (1999) *Mol. Cell.* 4: 735-743 discloses the use of the C/EBP tau activation domain, while the C/EBP alpha transactivation domain is disclosed in Tao, H., & Umek, RM. (1999) *DNA Cell Biol.* 18: 75-84.

It is known that zinc finger proteins may be fused to transcriptional repression domains such as the Kruppel-associated box (KRAB) domain to form powerful repressors. These fusions are known to repress expression of a reporter gene even when bound to sites a few kilobase pairs upstream from the promoter of the gene (Margolin *et al.*, 1994, *Proc. Natl. Acad. Sci. USA* 91: 4509-4513). In one preferred embodiment, the KRAB repressor domain from the human KOX-1 protein is used to repress gene activity (Moosmann *et al.*, *Biol. Chem.* 378: 669-677 (1997); Thiesen *et al.*, *New Biologist* 2: 363-374 (1990)). Other preferred transcriptional repressor domains are known in the art and include, for example, the *engrailed* domain (Han *et al.*, *EMBO J.* 12: 2723-2733 (1993)) and the *snag* domain (Grimes *et al.*, *Mol Cell. Biol.* 16: 6263-6272 (1996)). These can be used alone or in combination to down-regulate gene expression in animals.

Biological effector domains may be covalently or non-covalently linked to the nucleotide-binding domain. In a preferred embodiment the covalent linker comprises a amino acid sequence which may be flexible; polypeptides according to this embodiment preferably comprise fusion proteins comprising the nucleic acid binding portion of the chimera fused with an amino acid linker to the biological effector domain portion.
Alternatively, the covalent linker may comprise a synthetic, non-amino acid based, chemical linker, for example, polyethylene glycol. Synthetic linkers are commercially available, and methods of chemical conjugation are known in the art. The covalent linkers may comprise flexible or structured linkers, as described in detail above.

Non-covalent linkages between the nucleic acid binding portion and the effector portion may for example be formed using leucine zipper / coiled coil domains, or other naturally occurring or synthetic dimerisation domains (see e.g. Luscher, B. & Larsson, L. G. *Oncogene* 18:2955-2966 (1999) and Gouldson, P. R. *et al.*, *Neuropsychopharmacology* 23: S60-S77 (2000)).

TELOMERES, G-QUADRUPLLEXES AND G-QUARTETS

Telomeres comprise highly conserved DNA repeat sequences, associated with proteins, found at the ends of chromosomes in nearly all eukaryotes. They are widely studied because of their important roles in maintaining chromosome stability and in mediating normal chromosome segregation in mitosis and meiosis (Rhodes, D., & Giraldo, R. (1995) *Curr Opin Str Biol* 5, 311-322.).

Because of their potential importance, G-quadruplexes have been extensively characterised in terms of structure, polymorphism, ion selectivity, stability and folding kinetics [reviewed in (Williamson, J. R. (1994) *Annual Review Of Biophysics and Biomolecular Structure* 23, 703-730.)].

Telomeric DNA sequences contain characteristic guanine-rich repeats of the form $d(T_1-3-(T/A)-G_3-4)_n$ [reviewed in (Blackburn, E. H., & Szostak, J. W. (1984) *Annual*

Review Of Biochemistry 53, 163-194.)). These sequences form G-quadruplex secondary structures *in vitro* at physiological salt concentrations (K^+ or Na^+) and it has been proposed that such structures may be of biological significance *in vivo*. It has been suggested that inter-telomeric G-quadruplexes may determine the correct association of homologous chromosomes in different stages of the cell cycle (Sen, D., & Gilbert, W. (1988) *Nature* 334, 364-366. Sundquist, W. I., & Klug, A. (1989) *Nature* 342, 825-829. Williamson, J. R., Raghuraman, M. K., & Cech, T. R. (1989) *Cell* 59, 871-880.). More recently, it has been suggested that the G-quadruplex conformation of single stranded telomeric DNA may be important to the mechanism and regulation of telomerase-mediated telomere extension (Salazar M, Thompson BD, Kerwin SM, Hurley LH. (1996) *Biochemistry* 35(50):16110-5. Sun, D., LopezGuajardo, C. C., Quada, J., Hurley, L. H., & VonHoff, D. D. (1999) *Biochemistry* 38, 4037-4044.). Furthermore, G-quadruplexes have emerged as a molecular target for therapeutics particularly in cancer research (Sun D, Thompson B, Cathers BE, Salazar M, Kerwin SM, Trent JO, Jenkins TC, Neidle S, Hurley LH (1997) *J Med Chem* Jul 4;40(14):2113-6. Perry PJ, Reszka AP, Wood AA, Read MA, Gowan SM, Dosanjh HS, Trent JO, Jenkins TC, Kelland LR, Neidle S. (1998) *J Med Chem.* 41(24):4873-84.).

Several naturally occurring proteins with affinity for G-quadruplexes have been reported (Wellinger, R. J., & Sen, D. (1997) *European Journal of Cancer* 33, 735-749.), although there are problems associated with their use as diagnostic or therapeutic probes. Most examples, such as a recently reported DNA-binding autoantibody (Brown, B. A., Li, Y. Q., Brown, J. C., Hardin, C. C., Roberts, J. F., Pelsue, S. C., & Shultz, L. D. (1998) *Biochemistry* 37, 16325-16337.), have only moderate binding affinities and discriminate weakly between duplex and quadruplex DNA. Naturally occurring high-affinity telomere-binding proteins also appear unable to discriminate these structures. For example, *Saccharomyces cerevisiae* RAP1 (Giraldo, R., & Rhodes, D. (1994) *EMBO J* 13, 2411-2420.) has distinct but inseparable domains for binding quadruplexes and double stranded DNA.

Other known telomere-binding proteins have only moderate binding affinities and/or discriminate weakly between duplex and quadruplex DNA. The present invention makes use of the properties of enhanced binding affinity and ability to discriminate between duplex and quadruplex DNA of the nucleic acid binding polypeptides described
5 here.

Such nucleic acid binding polypeptides may therefore be used as probes for the presence of G-quadruplex structures, both *in vitro* and *in vivo*. ELISA-based detection of telomerase activity is therefore enabled. This detection system is rapid, easily automated with liquid handling robotics and avoids the need to use radioactivity. This contrasts with
10 prior art telomerase assays such as the commercially-available 'TRAP' assay.

Telomere-binding molecules described here may be used to target chromosome ends and deliver effector activity, for example using fusions with other peptides or enzymes.

It is envisaged that the present invention may be of use in diverse areas, including
15 for example one or more of the following; diagnostics, assays, ELISA testing, probe production, genomics studies such as pharmacogenomics, therapeutic applications such as study or construction of disease model(s), drug design, peptide/protein research, the construction or exploitation of research tools such as molecular marker(s) and similar reagents, as well as use in screening such as using chip technology, cellular or *in vitro*
20 assay(s), molecular detection as well as target identification or validation.

The nucleic acid binding polypeptides described here may also be of use in the study and/or treatment of metabolic disorders, or cancer. The present invention facilitates the construction of ELISA-based diagnostic kits for the detection of telomerase activity. These assays are rapid, easily automated with liquid handling robotics and avoid the need
25 to use radioactivity, in contrast to prior art technologies such as the 'TRAP' assay.

The present invention encompasses the development of probes for examining G-quartet formation *in vivo*. The nucleic acid binding polypeptides may also be used to detect

telomeric structures in any system, such as a cell or tissue or organ. The relevant system is exposed to a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid, and the binding between the nucleic acid binding polypeptide and any telomeric structures in the system is detected. Such
5 detection may employ any suitable means known in the art; furthermore, the nucleic acid binding polypeptide may preferably be labelled for this purpose. Labels are known in the art, and various ones may be used. The method set out above is suitable for localising telomeric structures, e.g., telomeres, in the system, and advantageously comprises detection of binding *in vivo* or *in situ*. The system preferably comprises a cell.

10 Telomere-binding molecules described here may be used to target chromosome ends and to deliver effector activity in the form of fusions with other peptides or enzymes. Therapeutic applications of the invention include those associated with the role(s) of telomeres in ageing and/or cancer.

15 Telomere-binding molecules described here may affect telomerase activity and may be used as, or in conjunction with, inhibitors of this enzyme, the activity of which is associated with cell immortalisation and cancer.

20 The nucleic acid binding polypeptides described here may be selected from a phage display library to bind G-quadruplex DNA structures of single stranded human telomeric sequences with selectivity and high affinity. Advantageously, these zinc fingers have no detectable affinity for a duplex DNA made up of the Htelo sequence and its
25 complementary strand. These molecules represent a new class of DNA-binding zinc fingers and have utility for both study and exploration of the molecules themselves, and of therapeutics and assays, in addition to their utility as *in vitro* or *in vivo* molecular probes to explore possible mechanisms of inhibition and regulation of telomerase-mediated telomere extension. The widespread conservation of G-quadruplex-forming sequences at
chromosome ends means that the molecules described here will find utility in a wide range of biological systems.

Since *in vitro* diagnostic methods for detecting G-quadruplexes, such as circular dichroism (Giraldo, R., Suzuki, M., Chapman, L., & Rhodes, D. (1994) *Proc Natl Acad Sci* 91, 7658-7662.) and dimethyl sulphate protection (Sen, D., & Gilbert, W. (1992) *Methods In Enzymology* 211, 191-199.), cannot be carried out in living cells, the invention
5 is useful in relation to derivatives (e.g. fluorescent derivatives) of these zinc fingers which may reveal the presence, location and relevance of these telomeric structures *in vivo*.

Molecules according to the present invention are useful in the binding of non-conventional nucleic acid structures. Examples of such structures include non-Watson-Crick base paired DNA, for example Hoogsteen base paired DNA or other variants.
10 Furthermore, non-conventional DNA structures include non-double helical DNA conformations.

Possible molecular mechanisms of inhibition are discussed, together with the potential for using engineered zinc fingers to interfere with the cellular processes associated with telomere function.

15 ENZYMATIC INHIBITION

The nucleic acid binding polypeptides disclosed here are capable of binding to telomeric, G-quadruplex, or G-quartet nucleic acid.

The nucleic acid binding polypeptides are further capable of inhibiting various enzymatic activities, including but not limited to telomerase activity, polymerase activity,
20 nucleic acid repair activity, repair activity, endonuclease activity, exonuclease activity, terminal transferase activity, primase activity, processivity activity, PCNA activity, replication activity, initiation activity, elongation activity, termination activity, licensing activity, etc. Preferably, the above activities are those that relate to DNA as a nucleic acid, including for example, DNA polymerase activity, DNA repair activity, DNA replication,
25 etc.

Preferably, the enzymatic activity comprises a polymerase activity, more preferably a DNA polymerase activity. Inhibition of polymerase activity is demonstrated in Example 4. The nucleic acid binding polypeptides described here are also capable of inhibiting telomerase activity, as demonstrated in Example 4. Preferably, the inhibition of telomerase activity is due to interference with the enzymological processing of telomeric DNA through the capacity of the nucleic acid binding polypeptide to bind G-quadruplex DNA.

The telomerase inhibiting activity of the nucleic acid binding polypeptides described here may be used in a number of ways. As it is known that telomerase activity is associated with immortalisation of cells, the nucleic acid binding polypeptides described here may be used to prevent such immortalisation. They can also be used to prevent proliferation of cells, or to induce differentiation of cells. The nucleic acid binding polypeptides described here also comprise anti-cancer and anti-tumour activity due to their ability to prevent cell proliferation. Accordingly, they may be used to treat cancer or prevent tumour formation, whether alone or in the form of pharmaceutical compositions optionally comprising other anti-cancer agents. Furthermore, other proliferative diseases are known in the art, and are suitably treated or prevented by means of the nucleic acid binding polypeptides described here. Examples of proliferative diseases and hyperproliferative diseases include skin proliferative diseases and inflammatory diseases such as psoriasis, dermatitis, etc.

The nucleic acid binding polypeptides described here also have other enzymatic inhibitory properties; for example, they may be used to inhibit activities of viral enzymes such as gp120 and integrase, preferably HIV gp120 and HIV integrase. These activities enable the use of the nucleic acid binding polypeptides described here for anti-viral purposes. Thus, the nucleic acid binding polypeptides are capable of preventing replication of a retrovirus such as HIV, when the virus, or a nucleic acid containing portion of the virus, is exposed to the nucleic acid binding polypeptide. A patient suffering from a viral or retroviral disease may be treated by administration of the nucleic acid binding polypeptide. The anti-viral properties of the nucleic acid binding polypeptides are further described elsewhere in this document.

In a highly preferred embodiment of the invention, the nucleic acid binding polypeptides described here are capable of inhibiting the enzymatic activity in the context of a cellular environment or *in vivo*. Thus, as shown in the Examples, the nucleic acid binding polypeptides are capable of binding to DNA within the environment of a cell, and inhibiting a cellular enzymatic activity. More preferably, the nucleic acid binding polypeptides are capable of inhibiting the enzymatic activity within a relevant cellular compartment, region or organelle; for example, inhibition of DNA replication activity within the nucleus.

ANTI-VIRAL PROPERTIES

The nucleic acid binding polypeptides described here may be used to prevent viral replication, and can therefore be used as treatments for viral diseases. Such a use takes advantage of their ability to bind viral nucleic acids which have or mimic quadruplex structures. Binding of the nucleic acid binding polypeptides to such structures prevents viral replication.

Without seeming to be bound by a particular theory, the nucleic acid binding polypeptides described here are believed to function by competing with a natural quadruplex target of essential HIV proteins such as HIV integrase and gp120, thereby inhibiting a required interaction. Thus, they are believed to function in a similar manner to other quadruplex binding drugs such as Zintevir (Cherepanov et al., 1997, . Mol Pharmacol 1997 Nov;52(5):771-80).

Accordingly, we further provide the use of the nucleic acid binding polypeptides described here as enzymatic inhibitors of integrase and gp120 activity.

We therefore provide further a method of targeting a native viral nucleic acid sequence with a nucleic acid binding polypeptide. The method comprises providing a nucleic acid binding polypeptide capable of binding to a telomeric, G-quadruplex, or G-quartet nucleic acid and also providing a native viral nucleic acid sequence comprising

one or more nucleotide sequences capable of being bound by the nucleic acid binding polypeptide. The nucleic acid binding polypeptide may be one which is capable of binding a structure in the nucleic acid sequence of the virus, such a structure mimicking a telomeric, G-quadruplex, or G-quartet structure. The nucleic acid binding polypeptide is then contacted with the native viral nucleic acid sequence. Preferably, the native viral nucleic acid mediates the infection of a cell by a virus. More preferably, the native viral nucleic acid sequence comprises a provirus or an virus integrated into the genome of a host cell.

We also provide a method of down-regulating a viral function in a cell infected with the virus, the method comprising contacting the virus and/or the cell with a nucleic acid binding polypeptide capable of binding a telomeric, G-quadruplex, or G-quartet nucleic acid of the virus, or a structure mimicking such a nucleic acid.

We further provide a method of modulating, preferably down-regulating, a viral function in a system comprising administering a polypeptide as described above to said system. Preferably, the viral function is selected from the group consisting of: viral titre, viral infectivity, viral replication, viral packaging, and viral transcription.

The nucleic acid binding polypeptides described here may be administered together with known anti-virals, for example, zinc finger polypeptides capable of binding to promoter or other regions of viral nucleotide sequences. Such zinc finger polypeptides are described in detail in our International Patent Application PCT/GB01/02017.

DRUG SCREENING

The nucleic acid binding polypeptides described here may also be used in drug screening applications. For example, they may be used to screen molecules, for example, small molecules for drug-like interactions. Examples of such molecules are those which bind to quadruplex nucleic acid.

Several ways of conducting such a screen may be envisaged. However, all of these rely on the ability of a suitable molecule to bind to a telomeric, G-quadruplex, or G-quartet nucleic acid. Such a binding may be detected by the disruptive effect of the molecule on the binding interaction between a telomeric, G-quadruplex, or G-quartet nucleic acid and a
5 nucleic acid binding polypeptide capable of binding to such a nucleic acid. Thus, a dissociation between the nucleic acid and the nucleic acid binding polypeptide may be detected in the presence of a suitable molecule. This may be done by detecting the binding or the strength of binding in the absence and presence of a suitable molecule.

The method may therefore comprise firstly providing a nucleic acid comprising a
10 telomeric, G-quadruplex, or G-quartet structure together with a nucleic acid binding polypeptide capable of binding to a nucleic acid comprising such a structure. Either or both of these are then contacted with a candidate molecule and binding between the nucleic acid and the nucleic acid binding polypeptide is determined.

The method may comprise monitoring the binding between a nucleic acid
15 comprising a telomeric, G-quadruplex, or G-quartet structure and a nucleic acid binding polypeptide capable of binding to a nucleic acid comprising such a structure, in the presence and absence of a candidate molecule.

The method may also comprise providing a complex between a nucleic acid comprising a telomeric, G-quadruplex, or G-quartet structure and a nucleic acid binding
20 polypeptide capable of binding to a nucleic acid comprising such a structure; contacting either or both members of the complex with a candidate molecule; and detecting a dissociation between the members of the complex.

The method preferably further comprises a step of isolating, synthesising and/or providing a composition comprising the candidate molecule identified to have such
25 activity.

The screen is preferably conducted in a high-throughput format, which enables a large number of candidate compounds to be screen simultaneously. Libraries of such compounds are commercially available a range of suppliers (eg Maybridge, UK), each with varying chemical synthesis strategies to cover 'library space.' Combinatorial libraries
5 may also be used. Preferably, arrays comprising such libraries are employed. Such arrays are described in detail elsewhere in this document.

For a screening approach to be successful, it is critical to develop a screening assay that is simple, reliable and suitable for automation. For example, ELISA or FRET-type binding assays are rapid and convenient, being easily adapted to high-throughput robotic
10 formats. Given a library of small molecules, it is therefore necessary to develop either some kind of tagging system to detect binding, or a specific functional assay. A preferred embodiment relies on the assaying of the disruption of specific, detectable complexes formed between the target and some kind of secondary marker molecule. Hence, by assaying the effect of the candidate drug on the target-marker interaction, a large number
15 of candidates can be rapidly and simply screened. According to this, the target comprises a quadruplex nucleic acid and the secondary marker molecule comprises a quadruplex binding polypeptide, for example, a zinc finger.

An exemplary screen to discover small molecule drugs that bind telomeric DNA employs the zinc finger protein Gq1 as a marker molecule. Since Gq1 binds specifically to
20 the human telomeric DNA repeats (with an affinity of ~30 nM) a drug screen may be carried out as shown in Figure 21.

The drug screen shown in Figure 21 depends on fluorescence resonance energy transfer (FRET). FRET is a distance-dependent interaction between the electronic excited states of two dye molecules in which excitation is transferred from a donor molecule to an
25 acceptor molecule *without emission of a photon*. FRET is dependent on the inverse sixth power of the intermolecular separation, making it useful over distances comparable with the dimensions of biological macromolecules. Thus, FRET is an important technique for investigating a variety of biological phenomena that produce changes in molecular

proximity. Any of the following dyes may be used in the Fluorescence Resonance Energy Transfer (FRET) reactions as described above.

Donor	Acceptor
Fluorescein	Tetramethylrhodamine
IAEDANS	Fluorescein
EDANS	DABCYL
Fluorescein	Fluorescein
BODIPY FL	BODIPY FL
Fluorescein	QSY 7 dye

FRET

The drug screening, telomerase assay, telomere length assay aspects of this invention make use of a step of determining the binding or dissociation between a nucleic acid binding polypeptide and a target nucleic acid. The determination may be made by various means, such as ELISA. Preferably, a signal which is the emission or absorption of electromagnetic radiation is detected. However, and preferably, the determination makes use of Fluorescence Resonance Energy Transfer (FRET).

FRET is detectable when two fluorescent labels which fluoresce at different frequencies are sufficiently close to each other that energy is able to be transferred from one label to the other. FRET is widely known in the art (for a review, see Matyus, 1992, *J. Photochem. Photobiol. B: Biol.*, 12: 323-337, which is herein incorporated by reference). FRET is a radiationless process in which energy is transferred from an excited donor

molecule to an acceptor molecule; the efficiency of this transfer is dependent upon the distance between the donor and acceptor molecules, as described below. Since the rate of energy transfer is inversely proportional to the sixth power of the distance between the donor and acceptor, the energy transfer efficiency is extremely sensitive to distance
5 changes. Energy transfer is said to occur with detectable efficiency in the 1-10 nm distance range, but is typically 4-6 nm for favourable pairs of donor and acceptor.

Accordingly, the invention may be practised by choosing suitable pairs of donor and acceptor molecules, and labelling the nucleic acid binding polypeptide and the nucleic acid target with these. When the two entities bind to each other, the donor molecule and
10 the acceptor molecule are brought together so that energy transfer occurs.

Radiationless energy transfer is based on the biophysical properties of fluorophores. These principles are reviewed elsewhere (Lakowicz, 1983, *Principles of Fluorescence Spectroscopy*, Plenum Press, New York; Jovin and Jovin, 1989, *Cell Structure and Function by Microspectrofluorometry*, eds. E. Kohen and J.G. Hirschberg,
15 Academic Press, both of which are incorporated herein by reference). Briefly, a fluorophore absorbs light energy at a characteristic wavelength. This wavelength is also known as the excitation wavelength. The energy absorbed by a fluorochrome is subsequently released through various pathways, one being emission of photons to produce fluorescence. The wavelength of light being emitted is known as the emission
20 wavelength and is an inherent characteristic of a particular fluorophore. Radiationless energy transfer is the quantum-mechanical process by which the energy of the excited state of one fluorophore is transferred without actual photon emission to a second fluorophore. That energy may then be subsequently released at the emission wavelength of the second fluorophore. The first fluorophore is generally termed the donor (D) and has an
25 excited state of higher energy than that of the second fluorophore, termed the acceptor (A). The essential features of the process are that the emission spectrum of the donor overlap with the excitation spectrum of the acceptor, and that the donor and acceptor be sufficiently close. The distance over which radiationless energy transfer is effective depends on many factors including the fluorescence quantum efficiency of the donor, the
30 extinction coefficient of the acceptor, the degree of overlap of their respective spectra, the

refractive index of the medium, and the relative orientation of the transition moments of the two fluorophores. In addition to having an optimum emission range overlapping the excitation wavelength of the other fluorophore, the distance between D and A must be sufficiently small to allow the radiationless transfer of energy between the fluorophores.

5 In a FRET assay, the fluorescent molecules are chosen such that the excitation spectrum of one of the molecules (the acceptor molecule) overlaps with the emission spectrum of the excited fluorescent molecule (the donor molecule). The donor molecule is excited by light of appropriate intensity within the donor's excitation spectrum. The donor then emits some of the absorbed energy as fluorescent light and dissipates some of the
10 energy by FRET to the acceptor fluorescent molecule. The fluorescent energy it produces is quenched by the acceptor fluorescent molecule. FRET can be manifested as a reduction in the intensity of the fluorescent signal from the donor, reduction in the lifetime of its excited state, and re-emission of fluorescent light at the longer wavelengths (lower energies) characteristic of the acceptor. When the donor and acceptor molecules become
15 spatially separated, FRET is diminished or eliminated.

Suitable fluorophores are known in the art, and include chemical fluorophores and fluorescent polypeptides, such as GFP and mutants thereof which fluoresce with different wavelengths or intensities (see WO 97/28261). Chemical fluorophores may be attached to immunoglobulin by incorporating binding sites therefor into the immunoglobulin during
20 the synthesis thereof.

Preferably, however, the fluorophore is a fluorescent protein, which is advantageously GFP or a mutant thereof. GFP and its mutants may be synthesised together with the binding means (where this is a polypeptide such as an immunoglobulin) by expression therewith as a fusion polypeptide, according to methods well known in the art.

ARRAYS

As described elsewhere, the nucleic acid binding polypeptide capable of binding to telomeric, G-quadruplex, or G-quartet nucleic acid may be used in a screen for molecules which interact with the quadruplex etc, or which disrupt the interaction between the quadruplex etc and the nucleic acid binding polypeptide. Such a screen may employ a library of candidate molecules, which library may preferably be in the form of an array.

Array technology and the various techniques and applications associated with it is described generally in numerous textbooks and documents. These include Lemieux et al., 1998, *Molecular Breeding* 4, 277-289, Schena and Davis. *Parallel Analysis with Biological Chips*. in *PCR Methods Manual* (eds. M. Innis, D. Gelfand, J. Sninsky), Schena and Davis, 1999, *Genes, Genomes and Chips*. In *DNA Microarrays: A Practical Approach* (ed. M. Schena), Oxford University Press, Oxford, UK, 1999), *The Chipping Forecast* (*Nature Genetics* special issue; January 1999 Supplement), Mark Schena (Ed.), *Microarray Biochip Technology*, (Eaton Publishing Company), Cortes, 2000, *The Scientist* 14[17]:25, Gwynne and Page, *Microarray analysis: the next revolution in molecular biology*, *Science*, 1999 August 6; Eakins and Chu, 1999, *Trends in Biotechnology*, 17, 217-218, and also at <http://www.gene-chips.com>.

Array technology overcomes the disadvantages with traditional methods in molecular biology, which generally work on a "one gene in one experiment" basis, resulting in low throughput and the inability to appreciate the "whole picture" of gene function. Currently, the major applications for array technology include the identification of sequence (gene / gene mutation) and the determination of expression level (abundance) of genes. Gene expression profiling may make use of array technology, optionally in combination with proteomics techniques (Celis et al, 2000, *FEBS Lett*, 480(1):2-16; Lockhart and Winzeler, 2000, *Nature* 405(6788):827-836; Khan et al., 1999, 20(2):223-9). Other applications of array technology are also known in the art; for example, gene discovery, cancer research (Marx, 2000, *Science* 289: 1670-1672; Scherf, et al, 2000, *Nat Genet*;24(3):236-44; Ross et al, 2000, *Nat Genet*. 2000 Mar;24(3):227-35), SNP analysis

(Wang et al, 1998, Science, 280(5366):1077-82), drug discovery, pharmacogenomics, disease diagnosis (for example, utilising microfluidics devices: Chemical & Engineering News, February 22, 1999, 77(8):27-36), toxicology (Rockett and Dix (2000), *Xenobiotica*, 30(2):155-77; Afshari et al., 1999, Cancer Res 59(19):4759-60) and toxicogenomics (a
5 hybrid of functional genomics and molecular toxicology). The goal of toxicogenomics is to find correlations between toxic responses to toxicants and changes in the genetic profiles of the objects exposed to such toxicants (Nuwaysir, et al (1999), *Molecular Carcinogenesis*, 24:153-159).

In general, any library may be arranged in an orderly manner into an array, by
10 spatially separating the members of the library. Examples of suitable libraries for arraying include nucleic acid libraries (including DNA, cDNA, oligonucleotide, etc libraries), peptide, polypeptide and protein libraries, as well as libraries comprising any molecules, such as ligand libraries, among others. Accordingly, where reference is made to a "library" in this document, unless the context dictates otherwise, such reference should be taken to
15 include reference to a library in the form of an array.

The samples (e.g., members of a library) are generally fixed or immobilised onto a solid phase, preferably a solid substrate, to limit diffusion and admixing of the samples. In a preferred embodiment, libraries of DNA binding ligands may be prepared. In particular, the libraries may be immobilised to a substantially planar solid phase, including
20 membranes and non-porous substrates such as plastic and glass. Furthermore, the samples are preferably arranged in such a way that indexing (i.e., reference or access to a particular sample) is facilitated. Typically the samples are applied as spots in a grid formation. Common assay systems may be adapted for this purpose. For example, an array may be immobilised on the surface of a microplate, either with multiple samples in a well, or with
25 a single sample in each well. Furthermore, the solid substrate may be a membrane, such as a nitrocellulose or nylon membrane (for example, membranes used in blotting experiments). Alternative substrates include glass, or silica based substrates. Thus, the samples are immobilised by any suitable method known in the art, for example, by charge interactions, or by chemical coupling to the walls or bottom of the wells, or the surface of
30 the membrane. Other means of arranging and fixing may be used, for example, pipetting,

drop-touch, piezoelectric means, ink-jet and bubblejet technology, electrostatic application, etc. In the case of silicon-based chips, photolithography may be utilised to arrange and fix the samples on the chip.

The samples may be arranged by being "spotted" onto the solid substrate; this may
5 be done by hand or by making use of robotics to deposit the sample. In general, arrays may be described as macroarrays or microarrays, the difference being the size of the sample spots. Macroarrays typically contain sample spot sizes of about 300 microns or larger and may be easily imaged by existing gel and blot scanners. The sample spot sizes in microarrays are typically less than 200 microns in diameter and these arrays usually
10 contain thousands of spots. Thus, microarrays may require specialized robotics and imaging equipment, which may need to be custom made. Instrumentation is described generally in a review by Cortese, 2000, *The Scientist* 14[11]:26.

Techniques for producing immobilised libraries of DNA molecules have been described in the art. Generally, most prior art methods described how to synthesise single-
15 stranded nucleic acid molecule libraries, using for example masking techniques to build up various permutations of sequences at the various discrete positions on the solid substrate. U.S. Patent No. 5,837,832, the contents of which are incorporated herein by reference, describes an improved method for producing DNA arrays immobilised to silicon substrates based on very large scale integration technology. In particular, U.S. Patent No.
20 5,837,832 describes a strategy called "tiling" to synthesize specific sets of probes at spatially-defined locations on a substrate which may be used to produced the immobilised DNA libraries of the present invention. U.S. Patent No. 5,837,832 also provides references for earlier techniques that may also be used.

Arrays of peptides (or peptidomimetics) may also be synthesised on a surface in a
25 manner that places each distinct library member (e.g., unique peptide sequence) at a discrete, predefined location in the array. The identity of each library member is determined by its spatial location in the array. The locations in the array where binding interactions between a predetermined molecule (e.g., a target or probe) and reactive library

members occur is determined, thereby identifying the sequences of the reactive library members on the basis of spatial location. These methods are described in U.S. Patent No. 5,143,854; WO90/15070 and WO92/10092; Fodor *et al.* (1991) *Science*, 251: 767; Dower and Fodor (1991) *Ann. Rep. Med. Chem.*, 26: 271.

5 To aid detection, targets and probes may be labelled with any readily detectable reporter, for example, a fluorescent, bioluminescent, phosphorescent, radioactive, etc reporter. Such reporters, their detection, coupling to targets/probes, etc are discussed elsewhere in this document. Labelling of probes and targets is also disclosed in Shalon *et al.*, 1996, *Genome Res* 6(7):639-45

10 Specific examples of DNA arrays are as follow:

Format I: probe cDNA (500~5,000 bases long) is immobilized to a solid surface such as glass using robot spotting and exposed to a set of targets either separately or in a mixture. This method is widely considered as having been developed at Stanford University (Ekins and Chu, 1999, *Trends in Biotechnology*, 1999, 17, 217-218).

15 Format II: an array of oligonucleotide (20~25-mer oligos) or peptide nucleic acid (PNA) probes is synthesized either in situ (on-chip) or by conventional synthesis followed by on-chip immobilization. The array is exposed to labeled sample DNA, hybridized, and the identity/abundance of complementary sequences are determined. Such a DNA chip is sold by Affymetrix, Inc., under the GeneChip® trademark.

20 Examples of some commercially available microarray formats are set out in Table 1 below (see also Marshall and Hodgson, 1998, *Nature Biotechnology*, 16(1), 27-31.

Company	Product name	Arraying method	Hybridization step	Readout
Affymetrix, Inc., Santa Clara, California	GeneChip®	<i>In situ</i> (on-chip) photolithographic synthesis of ~20-25-mer oligos onto silicon wafers, which are diced into 1.25 cm ² or 5.25 cm ² chips	10,000-260,000 oligo features probed with labeled 30-40 nucleotide fragments of sample cDNA or antisense RNA	Fluorescence

Brax, Cambridge, UK		Short synthetic oligo, synthesized off-chip	1000 oligos on a "universal chip" probed with tagged nucleic acid	Mass spectrometry
Gene Logic, Inc., Columbia, Maryland	READS ^T M			
Genometrix Inc., The Woodlands, Texas	Universal Arrays TM			
GENSET, Paris, France				
Hyseq Inc., Sunnyvale, California	HyChip TM	500-2000 nt DNA samples printed onto 0.6 cm ² (HyGnostics) or ~18 cm ² (Gene Discovery) membranes Fabricated 5-mer oligos printed as 1,15 cm ² arrays onto glass (HyChip)	64 sample cDNA spots probed with 8,000 7-mer oligos (HyGnostics) or <=55,000 sample cDNA spots probed with 300 7- mer oligo (Gene Discovery) Universal 1024 oligo spots probed 10 kb sample cDNAs, labeled 5-mer oligo, and ligase	Radioisotope Fluorescence
Incyte Pharmaceutica ls. Inc., Palo Alto, California	GEM	Piezoelectric printing for spotting PCR fragments and on-chip synthesis of oligos	<=1000 (eventually 10,000) oligo/PCR fragment spots probed with labeled RNA	Fluorescence and radioisotope
Molecular Dynamics, Inc., Sunnyvale, California	Storm® FluorIma ger®	500-5000 nt cDNAs printed by pen onto ~10 cm ² on glass slide	~10,000 cDNA spots probed with 200-400 nt labeled sample cDNAs	Fluorescence
Nanogen, San Diego, California	Semicond uctor Microchip	Prefabricated ~20-mer oligos, captured onto electroactive spots on silicon wafers, which are diced into <=1 cm ² chips	25, 64, 400 (and eventually 10,000) oligo spots polarized to enhance hybridization to 200-400 nt labeled sample cDNAs	Fluorescence
Protogene Laboratories, Palo Alto, California		On-chip synthesis of 40- 50-mer oligos onto 9 cm ² glass chip via printing to a surface- tension array	<=8,000 oligo spots probed with 200-400 nt labeled sample nucleic acids	Fluorescence
Sequenom, Hamburg, Germany, and San Diego, California	MassArra y SpectroC hip	Off-set printing of array; around 20-25-mer oligos	250 locations per SpectroChip interrogated by laser desorption and mass spectrometry	Mass spectrometry
Synteni, Inc., Fremont, California	UniGEM TM	500-5,000 nt cDNAs printed by tip onto ~4 cm ² glass chip	<=10,000 cDNA spots probed with 200-400 nt labeled sample cDNAs	Fluorescence
The German		Prototypic PNA	Around 1,000 spots on a	Fluorescence/mass

Cancer Institute, Heidelberg, Germany		macrochip with on-chip synthesis of probes using f-moc or t-moc chemistry	8 x 12 cm chip	spectrometry
--	--	--	----------------	--------------

Table 1. Examples of currently available hybridization microarray formats

Data analysis is also an important part of an experiment involving arrays. The raw data from a microarray experiment typically are images, which need to be transformed
 5 into gene expression matrices - tables where rows represent for example genes, columns represent for example various samples such as tissues or experimental conditions, and numbers in each cell for example characterize the expression level of the particular gene in the particular sample. These matrices have to be analyzed further, if any knowledge about the underlying biological processes is to be extracted. Methods of data analysis (including
 10 supervised and unsupervised data analysis as well as bioinformatics approaches) are disclosed in Brazma and Vilo J (2000) FEBS Lett 480(1):17-24.

As disclosed above, proteins, polypeptides, etc may also be immobilised in arrays. For example, antibodies have been used in microarray analysis of the proteome using protein chips (Borrebaeck CA, 2000, Immunol Today 21(8):379-82). Polypeptide arrays
 15 are reviewed in, for example, MacBeath and Schreiber, 2000, *Science*, 289(5485): p. 1760-1763.

VARIANTS

The nucleic acid binding polypeptide molecule as provided by the present invention includes splice variants encoded by mRNA generated by alternative splicing of a
 20 primary transcript, amino acid mutants, glycosylation variants and other covalent derivatives of said molecule which retain the physiological and/or physical properties of said molecule, such as its nucleic acid binding activity. Exemplary derivatives include molecules wherein the protein of the invention is covalently modified by substitution, chemical, enzymatic, or other appropriate means with a moiety other than a naturally
 25 occurring amino acid. Such a moiety may be a detectable moiety such as an enzyme or a

radioisotope, or may be a molecule capable of facilitating crossing of cell membrane(s) etc.

Derivatives can be fragments of the nucleic acid binding molecule. Fragments of said molecule comprise individual domains thereof, as well as smaller polypeptides
5 derived from the domains. Preferably, smaller polypeptides derived from the nucleic acid binding polypeptides described here define a single epitope which is characteristic of said molecule. Fragments may in theory be almost any size, as long as they retain one characteristic of the nucleic acid binding molecule. Preferably, fragments may be at least 3 amino acids and in length.

10 Derivatives of the nucleic acid binding molecule also comprise mutants thereof, which may contain amino acid deletions, additions or substitutions, subject to the requirement to maintain at least one feature characteristic of said molecule. Thus, conservative amino acid substitutions may be made substantially without altering the nature of the molecule, as may truncations from the N- or C- terminal ends, or the
15 corresponding 5'- or 3'- ends of a nucleic acid encoding it. Deletions or substitutions may moreover be made to the fragments of the molecule comprised by the invention. Nucleic acid binding molecule mutants may be produced from a DNA encoding a transcription protein which has been subjected to *in vitro* mutagenesis resulting e.g. in an addition, exchange and/or deletion of one or more amino acids. For example, substitutional,
20 deletional or insertional variants of the molecule can be prepared by recombinant methods and screened for nucleic acid binding activity as described herein.

The fragments, mutants and other derivatives of the polypeptide nucleic acid binding molecule preferably retain substantial homology with said molecule. As used herein, "homology" means that the two entities share sufficient characteristics for the
25 skilled person to determine that they are similar in origin and/or function. Preferably, homology is used to refer to sequence identity. Thus, the derivatives of the molecule preferably retain substantial sequence identity with the sequence of said molecule.

“Substantial homology”, where homology indicates sequence identity, means more than 75% sequence identity and most preferably a sequence identity of 90% or more. Amino acid sequence identity may be assessed by any suitable means, including the BLAST comparison technique which is well known in the art, and is described in Ausubel
5 *et al.*, Short Protocols in Molecular Biology (1999) 4th Ed, John Wiley & Sons, Inc.

MUTATIONS

Mutations may be performed by any method known to those of skill in the art. Preferred, however, is site-directed mutagenesis of a nucleic acid sequence encoding the protein of interest. A number of methods for site-directed mutagenesis are known in the
10 art, from methods employing single-stranded phage such as M13 to PCR-based techniques (see “PCR Protocols: A guide to methods and applications”, M.A. Innis, D.H. Gelfand, J.J. Sninsky, T.J. White (eds.). Academic Press, New York, 1990). Preferably, the commercially available Altered Site II Mutagenesis System (Promega) may be employed, according to the directions given by the manufacturer.

15 Screening of the proteins produced by mutant genes is preferably performed by expressing the genes and assaying the binding ability of the protein product. A simple and advantageously rapid method by which this may be accomplished is by phage display, in which the mutant polypeptides are expressed as fusion proteins with the coat proteins of filamentous bacteriophage, such as the minor coat protein pII of bacteriophage m13 or
20 gene III of bacteriophage Fd, and displayed on the capsid of bacteriophage transformed with the mutant genes. The target nucleic acid sequence is used as a probe to bind directly to the protein on the phage surface and select the phage possessing advantageous mutants, by affinity purification. The phage are then amplified by passage through a bacterial host, and subjected to further rounds of selection and amplification in order to enrich the mutant
25 pool for the desired phage and eventually isolate the preferred clone(s). Detailed methodology for phage display is known in the art and set forth, for example, in US Patent 5,223,409; Choo and Klug, (1995) Current Opinions in Biotechnology 6:431-436; Smith, (1985) Science 228:1315-1317; and McCafferty *et al.*, (1990) Nature 348:552-554; all

incorporated herein by reference. Vector systems and kits for phage display are available commercially, for example from Pharmacia.

The present invention allows the production of what are essentially artificial nucleic acid binding proteins. In these proteins, artificial analogues of amino acids may be used, to impart the proteins with desired properties or for other reasons. Thus, the term "amino acid", particularly in the context where "any amino acid" is referred to, means any sort of natural or artificial amino acid or amino acid analogue that may be employed in protein construction according to methods known in the art. Moreover, any specific amino acid referred to herein may be replaced by a functional analogue thereof, particularly an artificial functional analogue. The nomenclature used herein therefore specifically comprises within its scope functional analogues of the defined amino acids.

The nucleic acid binding polypeptides of use here are preferably zinc finger polypeptides. In other words, they comprise a Cys2-His2 zinc finger motif.

VECTORS

The nucleic acid encoding the nucleic acid binding polypeptides described here can be incorporated into vectors for further manipulation. As used herein, vector (or plasmid) refers to discrete elements that are used to introduce heterologous nucleic acid into cells for either expression or replication thereof. Selection and use of such vehicles are well within the skill of the person of ordinary skill in the art. Many vectors are available, and selection of appropriate vector will depend on the intended use of the vector, i.e. whether it is to be used for DNA amplification or for nucleic acid expression, the size of the DNA to be inserted into the vector, and the host cell to be transformed with the vector. Each vector contains various components depending on its function (amplification of DNA or expression of DNA) and the host cell for which it is compatible. The vector components generally include, but are not limited to, one or more of the following: an origin of replication, one or more marker genes, an enhancer element, a promoter, a transcription termination sequence and a signal sequence.

Both expression and cloning vectors generally contain nucleic acid sequence that enable the vector to replicate in one or more selected host cells. Typically in cloning vectors, this sequence is one that enables the vector to replicate independently of the host chromosomal DNA, and includes origins of replication or autonomously replicating sequences. Such sequences are well known for a variety of bacteria, yeast and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2 μ plasmid origin is suitable for yeast, and various viral origins (e.g. SV 40, polyoma, adenovirus) are useful for cloning vectors in mammalian cells. Generally, the origin of replication component is not needed for mammalian expression vectors unless these are used in mammalian cells competent for high level DNA replication, such as COS cells.

Most expression vectors are shuttle vectors, i.e. they are capable of replication in at least one class of organisms but can be transfected into another class of organisms for expression. For example, a vector is cloned in *E. coli* and then the same vector is transfected into yeast or mammalian cells even though it is not capable of replicating independently of the host cell chromosome. DNA may also be replicated by insertion into the host genome. However, the recovery of genomic DNA encoding the nucleic acid binding protein is more complex than that of exogenously replicated vector because restriction enzyme digestion is required to excise nucleic acid binding protein DNA. DNA can be amplified by PCR and be directly transfected into the host cells without any replication component.

SELECTABLE MARKERS

Advantageously, an expression and cloning vector may contain a selection gene also referred to as selectable marker. This gene encodes a protein necessary for the survival or growth of transformed host cells grown in a selective culture medium. Host cells not transformed with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that confer resistance to antibiotics and other toxins, e.g. ampicillin, neomycin, methotrexate or tetracycline,

complement auxotrophic deficiencies, or supply critical nutrients not available from complex media.

As to a selective gene marker appropriate for yeast, any marker gene can be used which facilitates the selection for transformants due to the phenotypic expression of the marker gene. Suitable markers for yeast are, for example, those conferring resistance to
5 antibiotics G418, hygromycin or bleomycin, or provide for prototrophy in an auxotrophic yeast mutant, for example the URA3, LEU2, LYS2, TRP1, or HIS3 gene.

Since the replication of vectors is conveniently done in *E. coli*, an *E. coli* genetic marker and an *E. coli* origin of replication are advantageously included. These can be
10 obtained from *E. coli* plasmids, such as pBR322, Bluescript® vector or a pUC plasmid, e.g. pUC18 or pUC19, which contain both *E. coli* replication origin and *E. coli* genetic marker conferring resistance to antibiotics, such as ampicillin.

Suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up nucleic acid binding protein nucleic acid, such
15 as dihydrofolate reductase (DHFR, methotrexate resistance), thymidine kinase, or genes conferring resistance to G418 or hygromycin. The mammalian cell transformants are placed under selection pressure which only those transformants which have taken up and are expressing the marker are uniquely adapted to survive. In the case of a DHFR or glutamine synthase (GS) marker, selection pressure can be imposed by culturing the
20 transformants under conditions in which the pressure is progressively increased, thereby leading to amplification (at its chromosomal integration site) of both the selection gene and the linked DNA that encodes the nucleic acid binding protein. Amplification is the process by which genes in greater demand for the production of a protein critical for growth, together with closely associated genes which may encode a desired protein, are
25 reiterated in tandem within the chromosomes of recombinant cells. Increased quantities of desired protein are usually synthesised from thus amplified DNA.

EXPRESSION

Expression and cloning vectors usually contain a promoter that is recognised by the host organism and is operably linked to nucleic acid binding protein encoding nucleic acid. Such a promoter may be inducible or constitutive. The promoters are operably linked
5 to DNA encoding the nucleic acid binding protein by removing the promoter from the source DNA by restriction enzyme digestion and inserting the isolated promoter sequence into the vector. Both the native nucleic acid binding protein promoter sequence and many heterologous promoters may be used to direct amplification and/or expression of nucleic acid binding protein encoding DNA

10 Promoters suitable for use with prokaryotic hosts include, for example, the β -lactamase and lactose promoter systems, alkaline phosphatase, the tryptophan (Trp) promoter system and hybrid promoters such as the tac promoter. Their nucleotide sequences have been published, thereby enabling the skilled worker operably to ligate them to DNA encoding nucleic acid binding protein, using linkers or adapters to supply
15 any required restriction sites. Promoters for use in bacterial systems will also generally contain a Shine-Delgarno sequence operably linked to the DNA encoding the nucleic acid binding protein.

Preferred expression vectors are bacterial expression vectors which comprise a promoter of a bacteriophage such as phage ϕ or T7 which is capable of functioning in the
20 bacteria. In one of the most widely used expression systems, the nucleic acid encoding the fusion protein may be transcribed from the vector by T7 RNA polymerase (Studier et al, Methods in Enzymol. 185; 60-89, 1990). In the *E. coli* BL21(DE3) host strain, used in conjunction with pET vectors, the T7 RNA polymerase is produced from the λ -lysogen DE3 in the host bacterium, and its expression is under the control of the IPTG inducible
25 lac UV5 promoter. This system has been employed successfully for over-production of many proteins. Alternatively the polymerase gene may be introduced on a lambda phage by infection with an int- phage such as the CE6 phage which is commercially available (Novagen, Madison, USA). other vectors include vectors containing the lambda PL

promoter such as PLEX (Invitrogen, NL), vectors containing the trc promoters such as pTrcHisXpressTm (Invitrogen) or pTrc99 (Pharmacia Biotech, SE) or vectors containing the tac promoter such as pKK223-3 (Pharmacia Biotech) or PMAL (New England Biolabs, MA, USA).

- 5 Moreover, the nucleic acid binding proteins described here preferably include a secretion sequence in order to facilitate secretion of the polypeptide from bacterial hosts, such that it will be produced as a soluble native peptide rather than in an inclusion body. The peptide may be recovered from the bacterial periplasmic space, or the culture medium, as appropriate. A "leader" peptide may be added to the N-terminal finger.
- 10 Preferably, the leader peptide is MAEEKP.

- Suitable promoting sequences for use with yeast hosts may be regulated or constitutive and are preferably derived from a highly expressed yeast gene, especially a *Saccharomyces cerevisiae* gene. Thus, the promoter of the TRP1 gene, the ADHI or ADHII gene, the acid phosphatase (PH05) gene, a promoter of the yeast mating
- 15 pheromone genes coding for the α - or α -factor or a promoter derived from a gene encoding a glycolytic enzyme such as the promoter of the enolase, glyceraldehyde-3-phosphate dehydrogenase (GAP), 3-phospho glycerate kinase (PGK), hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triose phosphate isomerase, phosphoglucose
- 20 isomerase or glucokinase genes, or a promoter from the TATA binding protein (TBP) gene can be used. Furthermore, it is possible to use hybrid promoters comprising upstream activation sequences (UAS) of one yeast gene and downstream promoter elements including a functional TATA box of another yeast gene, for example a hybrid promoter including the UAS(s) of the yeast PH05 gene and downstream promoter elements
- 25 including a functional TATA box of the yeast GAP gene (PH05-GAP hybrid promoter). A suitable constitutive PH05 promoter is e.g. a shortened acid phosphatase PH05 promoter devoid of the upstream regulatory elements (UAS) such as the PH05 (-173) promoter element starting at nucleotide -173 and ending at nucleotide -9 of the PH05 gene.

Nucleic acid binding protein gene transcription from vectors in mammalian hosts may be controlled by promoters derived from the genomes of viruses such as polyoma virus, adenovirus, fowlpox virus, bovine papilloma virus, avian sarcoma virus, cytomegalovirus (CMV), a retrovirus and Simian Virus 40 (SV40), from heterologous
5 mammalian promoters such as the actin promoter or a very strong promoter, e.g. a ribosomal protein promoter, and from the promoter normally associated with nucleic acid binding protein sequence, provided such promoters are compatible with the host cell systems.

Transcription of a DNA encoding nucleic acid binding protein by higher
10 eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are relatively orientation and position independent. Many enhancer sequences are known from mammalian genes (e.g. elastase and globin). However, typically one will employ an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270) and the CMV early promoter enhancer. The
15 enhancer may be spliced into the vector at a position 5' or 3' to nucleic acid binding protein DNA, but is preferably located at a site 5' from the promoter.

Advantageously, a eukaryotic expression vector encoding a nucleic acid binding protein described here may comprise a locus control region (LCR). LCRs are capable of directing high-level integration site independent expression of transgenes integrated into
20 host cell chromatin, which is of importance especially where the nucleic acid binding protein gene is to be expressed in the context of a permanently-transfected eukaryotic cell line in which chromosomal integration of the vector has occurred, or in transgenic animals.

Eukaryotic vectors may also contain sequences necessary for the termination of
25 transcription and for stabilising the mRNA. Such sequences are commonly available from the 5' and 3' untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding nucleic acid binding protein.

An expression vector includes any vector capable of expressing nucleic acid binding protein nucleic acids that are operatively linked with regulatory sequences, such as promoter regions, that are capable of expression of such DNAs. Thus, an expression vector refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector, that upon introduction into an appropriate host cell, results in expression of the cloned DNA. Appropriate expression vectors are well known to those with ordinary skill in the art and include those that are replicable in eukaryotic and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome. For example, DNAs encoding nucleic acid binding protein may be inserted into a vector suitable for expression of cDNAs in mammalian cells, e.g. a CMV enhancer-based vector such as pEVRF (Matthias, et al., (1989) NAR 17, 6418).

Particularly useful for practising the present invention are expression vectors that provide for the transient expression of DNA encoding nucleic acid binding protein in mammalian cells. Transient expression usually involves the use of an expression vector that is able to replicate efficiently in a host cell, such that the host cell accumulates many copies of the expression vector, and, in turn, synthesises high levels of nucleic acid binding protein. For the purposes of the present invention, transient expression systems are useful e.g. for identifying nucleic acid binding protein mutants, to identify potential phosphorylation sites, or to characterise functional domains of the protein.

Construction of vectors employs conventional ligation techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to generate the plasmids required. If desired, analysis to confirm correct sequences in the constructed plasmids is performed in a known fashion. Suitable methods for constructing expression vectors, preparing in vitro transcripts, introducing DNA into host cells, and performing analyses for assessing nucleic acid binding protein expression and function are known to those skilled in the art. Gene presence, amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA, dot blotting (DNA or RNA analysis), or in situ hybridisation, using an appropriately labelled probe which may be based on a

sequence provided herein. Those skilled in the art will readily envisage how these methods may be modified, if desired.

In accordance with another embodiment of the present invention, there are provided cells containing the above-described nucleic acids. Such host cells such as

5 prokaryote, yeast and higher eukaryote cells may be used for replicating DNA and producing the nucleic acid binding protein. Suitable prokaryotes include eubacteria, such as Gram-negative or Gram-positive organisms, such as *E. coli*, e.g. *E. coli* K-12 strains, DH5 α and HB101, or Bacilli. Further hosts suitable for the nucleic acid binding protein

10 encoding vectors include eukaryotic microbes such as filamentous fungi or yeast, e.g. *Saccharomyces cerevisiae*. Higher eukaryotic cells include insect and vertebrate cells, particularly mammalian cells including human cells or nucleated cells from other multicellular organisms. In recent years propagation of vertebrate cells in culture (tissue culture) has become a routine procedure. Examples of useful mammalian host cell lines

15 are epithelial or fibroblastic cell lines such as Chinese hamster ovary (CHO) cells, NIH 3T3 cells, HeLa cells or 293T cells. The host cells referred to in this disclosure comprise cells in *in vitro* culture as well as cells that are within a host animal.

DNA may be stably incorporated into cells or may be transiently expressed using methods known in the art. Stably transfected mammalian cells may be prepared by

20 transfecting cells with an expression vector having a selectable marker gene, and growing the transfected cells under conditions selective for cells expressing the marker gene. To prepare transient transfectants, mammalian cells are transfected with a reporter gene to monitor transfection efficiency.

To produce such stably or transiently transfected cells, the cells should be transfected with a sufficient amount of the nucleic acid binding protein-encoding nucleic

25 acid to form the nucleic acid binding protein. The precise amounts of DNA encoding the nucleic acid binding protein may be empirically determined and optimised for a particular cell and assay.

Host cells are transfected or, preferably, transformed with the above-captioned expression or cloning vectors of this invention and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. Heterologous DNA may be introduced into host
5 cells by any method known in the art, such as transfection with a vector encoding a heterologous DNA by the calcium phosphate coprecipitation technique or by electroporation. Numerous methods of transfection are known to the skilled worker in the field. Successful transfection is generally recognised when any indication of the operation of this vector occurs in the host cell. Transformation is achieved using standard techniques
10 appropriate to the particular host cells used.

Incorporation of cloned DNA into a suitable expression vector, transfection of eukaryotic cells with a plasmid vector or a combination of plasmid vectors, each encoding one or more distinct genes or with linear DNA, and selection of transfected cells are well known in the art (see, e.g. Sambrook et al. (1989) Molecular Cloning: A Laboratory
15 Manual, Second Edition, Cold Spring Harbor Laboratory Press).

Transfected or transformed cells are cultured using media and culturing methods known in the art, preferably under conditions, whereby the nucleic acid binding protein encoded by the DNA is expressed. The composition of suitable media is known to those in the art, so that they can be readily prepared. Suitable culturing media are also
20 commercially available.

Nucleic acid binding molecules as described here may be employed in a wide variety of applications, including diagnostics and as research tools. Advantageously, they may be employed as diagnostic tools for identifying the presence of nucleic acid molecules in a complex mixture.

25 Zinc finger domains may be made by methods described and/or referred to herein. For example, said zinc finger DNA binding domains may be made as discussed in the

examples, or as described in one or more of WO96/06166, WO98/53058, WO98/53057, or WO/98/53060.

FUSIONS

According to a further aspect, the invention provides a nucleic acid binding
5 polypeptide capable of binding to telomeric, G-quadruplex, or G-quartet nucleic acid
wherein said polypeptide comprises a nucleic acid binding domain and one or more further
domain(s) joined thereto. Said domains may be joined by any suitable means known in the
art, such as by conjugation, fusion, or other suitable method. Preferably, said domains are
comprised by a single polypeptide fusion protein. Such a nucleic acid binding polypeptide
10 may comprise nucleic acid binding domains linked by at least one flexible linker, one or
more domains linked by at least one structured linker, or both.

According to a further aspect, the invention provides a nucleic acid binding
polypeptide comprising a repressor domain and one or more nucleic acid binding domains.
The repressor domain is preferably a transcriptional repressor domain selected from the
15 group consisting of: a KRAB-A domain, an *engrailed* domain and a *snag* domain.

The nucleic acid binding polypeptides according to our invention may be linked to
one or more transcriptional effector domains, such as an activation domain or a repressor
domain. Examples of transcriptional activation domains include the VP16 and VP64
transactivation domains of Herpes Simplex Virus. Alternative transactivation domains are
20 various and include the maize C1 transactivation domain sequence (Sainz *et al.*, 1997,
Mol. Cell. Biol. 17: 115-22) and P1 (Goff *et al.*, 1992, Genes Dev. 6: 864-75; Estruch *et al.*,
1994, Nucleic Acids Res. 22: 3983-89) and a number of other domains that have been
reported from plants (see Estruch *et al.*, 1994, *ibid*).

Instead of incorporating a transactivator of gene expression, a repressor of gene
25 expression can be fused to the nucleic acid binding polypeptide and used to down regulate
the expression of a gene contiguous or incorporating the nucleic acid binding polypeptide

target sequence. Such repressors are known in the art and include, for example, the KRAB-A domain (Moosmann *et al.*, Biol. Chem. 378: 669-677 (1997)), the KRAB domain from human KOX1 protein (Margolin *et al.*, PNAS 91:4509-4513 (1994)), the *engrailed* domain (Han *et al.*, Embo J. 12: 2723-2733 (1993)) and the *snag* domain (Grimes *et al.*, Mol Cell. Biol. 16: 6263-6272 (1996)). These can be used alone or in combination to down-regulate gene expression.

The zinc finger proteins may be fused to transcriptional repression domains such as the Kruppel-associated box (KRAB) domain to form powerful repressors. These fusions are known to repress expression of a reporter gene even when bound to sites a few kilobase pairs upstream from the promoter of the gene (Margolin *et al.*, 1994, PNAS USA 91, 4509-4513).

Nucleic acid binding molecules may comprise tag sequences to facilitate studies and/or preparation of such molecules. Tag sequences may include flag-tag, myc-tag, 6his-tag or any other suitable tag known in the art.

Advantageously, such nucleic acid binding polypeptides may be used in combination. Use in combination includes both fusion of molecules into a single polypeptide as well as use of two or more discrete polypeptide molecules in solution.

The invention thus relates to the manipulation of telomeric structures using zinc finger peptides and derivative fusion proteins. Examples of such manipulation include simple binding, modification eg. methylation, cleavage eg. by nuclease action, or other chemical or physical modification.

Further fusion proteins are described herein, for example in the following section.

PHARMACEUTICALS

Moreover, the invention provides therapeutic agents and methods of therapy involving use of nucleic acid binding proteins as described herein. In particular, the invention provides the use of polypeptide fusions comprising an integrase, such as a viral
5 integrase, and a nucleic acid binding protein to target nucleic acid sequences *in vivo* (Bushman, (1994) PNAS (USA) 91:9233-9237). In gene therapy applications, the method may be applied to the delivery of functional genes into defective genes, or the delivery of nonsense nucleic acid in order to disrupt undesired nucleic acid. Alternatively, genes may be delivered to known, repetitive stretches of nucleic acid, such as centromeres, together
10 with an activating sequence such as an LCR. This would represent a route to the safe and predictable incorporation of nucleic acid into the genome.

In conventional therapeutic applications, nucleic acid binding proteins as described heremay be used to specifically knock out cell having mutant vital proteins. For example, if cells with mutant ras are targeted, they will be destroyed because ras is essential to
15 cellular survival. Alternatively, the action of transcription factors may be modulated, preferably reduced, by administering to the cell agents which bind to the binding site specific for the transcription factor. For example, the activity of HIV tat may be reduced by binding proteins specific for HIV TAR.

Moreover, binding proteins may be coupled to toxic molecules, such as nucleases,
20 which are capable of causing irreversible nucleic acid damage and cell death. Such nucleases include restriction endonuclease domains, non-specific nucleases such as DNase, RNase or similar enzymatic activities. Such agents are capable of selectively destroying cells which comprise a mutation in their endogenous nucleic acid.

Nucleic acid binding proteins and derivatives thereof as set forth above may also
25 be applied to the treatment of infections and the like in the form of organism-specific antibiotic or antiviral drugs. In such applications, the binding proteins may be coupled to a

nuclease or other nuclear toxin and targeted specifically to the nucleic acids of microorganisms.

The invention likewise relates to pharmaceutical preparations which contain the compounds or pharmaceutically acceptable salts thereof as active ingredients, and to
5 processes for their preparation.

The pharmaceutical preparations which contain the compound or pharmaceutically acceptable salts thereof are those for enteral, such as oral, furthermore rectal, and parenteral administration to (a) warm-blooded animal(s), the pharmacological active ingredient being present on its own or together with a pharmaceutically acceptable carrier.
10 The daily dose of the active ingredient depends on the age and the individual condition and also on the manner of administration.

The novel pharmaceutical preparations contain, for example, from about 10 % to about 80%, preferably from about 20 % to about 60 %, of the active ingredient. Pharmaceutical preparations for enteral or parenteral administration are, for example,
15 those in unit dose forms, such as sugar-coated tablets, tablets, capsules or suppositories, and furthermore ampoules. These are prepared in a manner known per se, for example by means of conventional mixing, granulating, sugar-coating, dissolving or lyophilising processes. Thus, pharmaceutical preparations for oral use can be obtained by combining the active ingredient with solid carriers, if desired granulating a mixture obtained, and
20 processing the mixture or granules, if desired or necessary, after addition of suitable excipients to give tablets or sugar-coated tablet cores.

Suitable carriers are, in particular, fillers, such as sugars, for example lactose, sucrose, mannitol or sorbitol, cellulose preparations and/or calcium phosphates, for example tricalcium phosphate or calcium hydrogen phosphate, furthermore binders, such
25 as starch paste, using, for example, corn, wheat, rice or potato starch, gelatin, tragacanth, methylcellulose and/or polyvinylpyrrolidone, if desired, disintegrants, such as the abovementioned starches, furthermore carboxymethyl starch, crosslinked

- polyvinylpyrrolidone, agar, alginic acid or a salt thereof, such as sodium alginate; auxiliaries are primarily glidants, flow-regulators and lubricants, for example silicic acid, talc, stearic acid or salts thereof, such as magnesium or calcium stearate, and/or polyethylene glycol. Sugar-coated tablet cores are provided with suitable coatings which, if desired, are resistant to gastric juice, using, inter alia, concentrated sugar solutions which, if desired, contain gum arabic, talc, polyvinylpyrrolidone, polyethylene glycol and/or titanium dioxide, coating solutions in suitable organic solvents or solvent mixtures or, for the preparation of gastric juice-resistant coatings, solutions of suitable cellulose preparations, such as acetylcellulose phthalate or hydroxypropylmethylcellulose phthalate.
- 10 Colorants or pigments, for example to identify or to indicate different doses of active ingredient, may be added to the tablets or sugar-coated tablet coatings.

- Other orally utilisable pharmaceutical preparations are hard gelatin capsules, and also soft closed capsules made of gelatin and a plasticiser, such as glycerol or sorbitol. The hard gelatin capsules may contain the active ingredient in the form of granules, for example in a mixture with fillers, such as lactose, binders, such as starches, and/or lubricants, such as talc or magnesium stearate, and, if desired, stabilisers. In soft capsules, the active ingredient is preferably dissolved or suspended in suitable liquids, such as fatty oils, paraffin oil or liquid polyethylene glycols, it also being possible to add stabilisers.
- 15

- Suitable rectally utilisable pharmaceutical preparations are, for example, suppositories, which consist of a combination of the active ingredient with a suppository base. Suitable suppository bases are, for example, natural or synthetic triglycerides, paraffin hydrocarbons, polyethylene glycols or higher alkanols. Furthermore, gelatin rectal capsules which contain a combination of the active ingredient with a base substance may also be used. Suitable base substances are, for example, liquid triglycerides, polyethylene glycols or paraffin hydrocarbons.
- 20
25

Suitable preparations for parenteral administration are primarily aqueous solutions of an active ingredient in water-soluble form, for example a water-soluble salt, and furthermore suspensions of the active ingredient, such as appropriate oily injection

suspensions, using suitable lipophilic solvents or vehicles, such as fatty oils, for example sesame oil, or synthetic fatty acid esters, for example ethyl oleate or triglycerides, or aqueous injection suspensions which contain viscosity-increasing substances, for example sodium carboxymethylcellulose, sorbitol and/or dextran, and, if necessary, also stabilisers.

- 5 The dose of the active ingredient depends on the warm-blooded animal species, the age and the individual condition and on the manner of administration. In the normal case, an approximate daily dose of about 10 mg to about 250 mg is to be estimated in the case of oral administration for a patient weighing approximately 75 kg.

- 10 An oligonucleotide drug (Zintevir; Aronex pharmaceuticals), that forms G-quadruplexes, inhibits HIV integrase and gp120 has already undergone phase I/II clinical trials (International Antiviral News 9:1, 2001). The oligo functions by competing with a natural quadruplex target of these HIV proteins, thereby inhibiting a required interaction (Mode of interaction of G-quartets with the integrase of human immunodeficiency virus type 1. Cherepanov P, Este JA, Rando RF, Ojwang JO, Reekmans G, Steinfeld
15 R, David G, De Clercq E, Debyser Z. Mol Pharmacol 1997 Nov;52(5):771-80).

Thus, we provide that zinc fingers which bind G-quadruplexes similarly prevent these HIV-proteins from interacting correctly with their natural substrates. Therefore Gq1 and derivative peptides provide a new class of anti-HIV agents.

EXAMPLES

20 **Example 1: Production of Molecules Binding G-Quadruplex Structures**

In this Example, DNA-binding proteins of the zinc finger family are engineered to bind specifically to a telomeric G-quadruplex nucleic acid structure.

A zinc finger library is screened for molecules that bind to an oligonucleotide containing the human telomeric repeat sequence in the G-quadruplex conformation. The

selected molecular clones exhibit amino acid homologies (consensus sequences). Without wishing to be bound by theory, this suggests that the molecules have analogous modes of binding. Binding is both sequence-dependent and structure-specific. This is the first example of a designed molecule that binds to G-quadruplex DNA. Further, this represents
5 a new type of binding interaction for a zinc finger protein molecule.

G-quadruplex DNA Ligand preparation

It has been previously reported that the human telomeric sequence (5'-GTTAGG-3')_n forms G-quadruplex structures *in vitro* (Balagurumoorthy, P.; Brahmachari, S. K., Mohanty, D., Bansal, M., & Sasisekharan, V. (1992) *Nucleic Acids Research* 20, 4061-4067. Balagurumoorthy, P., & Brahmachari, S. K. (1994) *Journal Of Biological Chemistry* 269, 21858-21869. Fletcher, T. M., Sun, D. K., Salazar, M., & Hurley, L. H. (1998) *Biochemistry* 37, 5536-5541.). The five repeat telomeric oligonucleotide sequence (5'-GTTAGG-3')₅ (Htelo) is employed as the ligand for affinity selection of phage herein.

Synthesised oligonucleotides (Oswel Ltd.) are purified by fractionation in
15 denaturing polyacrylamide-urea gels, recovered by elution and desalted further using Waters sep-Pack C-18 cartridges with final elution in 25 mM Tris, pH 7.5 as described by Giraldo et al. (Giraldo, R., & Rhodes, D. (1994) *EMBO J* 13, 2411-2420.).

The sequence 5'-biotin-GGTTAG GGTTAG GGTTAG GGTTAG GGTTAG-3' ('Biotin-Htelo') is prepared for the phage selection experiments and the unbiotinylated
20 sequence ('Htelo') is used for gel mobility shift and DMS protection experiments.

Oligonucleotides are then annealed for quadruplex formation, and subsequently used for ELISA and/or gel assays (see below). Oligonucleotides are diluted to 10 pmol/μl in 25 mM Tris (pH 7.5) or phosphate-buffered KCl or NaCl (pH 7.5) with cation concentrations ranging from 25 mM to 150 mM. Annealing or quadruplex formation is
25 carried out by heating samples to 95°C on a thermal heating block, and cooling to 4°C at a

rate of 2°C/min. The double stranded DNA (ds Htelo) is made by primer extension with the *Klenow* fragment of DNA polymerase.

Structures formed by human telomeric sequences may be analysed using dimethyl sulphate protection analysis to determine the existence of G-quadruplexes therein. To confirm that Htelo is folded into a G-quadruplex in the presence of sodium and potassium ions, a dimethyl sulphate (DMS) protection assay is carried out (Sundquist, W. I., & Klug, A. (1989) *Nature* 342, 825-829.). G-quadruplex formation involves Hoogsteen-type base pairing of guanines which protects N-7 of guanine against methylation on exposure to the potent methylating agent DMS. Subsequent cleavage of the DNA backbone at methylated guanines can be mediated by heating in aqueous piperidine (Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.).

The resulting gel pattern (see for example Figure 2A) clearly shows that the critical guanines of Htelo are almost completely protected from cleavage, at K⁺ or Na⁺ concentrations above 100 mM, as compared to a Tris-HCl buffer control. Non-denaturing gels confirm that these folded forms are of a single species and therefore antiparallel intramolecular G-quadruplexes, ie. similar to the structure illustrated in Figure 2B. Intermolecular G-quadruplexes are not observed in detectable amounts under these conditions. Without wishing to be bound by theory, this is probably because of their slow folding kinetics and/or because of the relatively low concentrations of DNA used which are likely to promote intramolecular G-quadruplex formation (Hardin CC, Henderson E, Watson T, Prosser JK (1991) *Biochemistry* May 7;30(18):4460-72).

A zinc finger phage display library is constructed specifically to select candidates that bind human telomeric DNA sequences, under conditions that promote G-quadruplex formation. The library is made up of zinc fingers with selectively randomised residues, biased for dsDNA binding potential (Choo, Y., & Klug, A. (1994) *Proc. Natl. Acad. Sci. U.S.A.* 91, 11163-11167. Isalan, M., Klug, A., & Choo, Y. (1998) *Biochemistry* 37, 12026-12033). Similar libraries have been extensively characterised, both biochemically and structurally, but only in their capacity to bind duplex DNA sequences in the major groove.

(Choo, Y., & Klug, A. (1997) *Curr. Opin. Str. Biol.* 7, 117-125. Choo, Y., & Isalan, M. D. (2000) *Current Opinion in Structural Biology* 10).

Because of practicalities of library handling, a complementary sub-library strategy is employed. Consequently, two complete sub-libraries are constructed and enriched for
5 DNA-binding potential by selection against randomised dsDNA sequences (see below). The resulting clones are recombined *in vitro* to make a library containing randomisations over all three fingers.

Construction of phage display library

A phage display library is constructed, based on the three-finger DNA-binding
10 domain of Zif268, whose structure is well characterised (Elrod-Erickson, M., Rould, M. A., Neklodova, L., & Pabo, C. O. (1996) *Structure* 4, 1171-1180. Pavletich, N. P., & Pabo, C. O. (1991) *Science* 252, 809-817.).

A zinc finger DNA-binding domain library is constructed comprising the amino acid framework of wild-type Zif268, but containing randomisations in amino acid
15 positions over all three fingers (see Figure1). Due to the practicalities of library cloning (ie. working with about $\sim 10^6$ - 10^7 transformants), the final library is advantageously constructed from two complementary sub-libraries: Sub-library-1 contains randomisations in F1 (-1 \rightarrow 6) and F2 (-1 \rightarrow 3). Conversely, sub-library-2 contains randomisations in F2 (3 \rightarrow 6) and F3 (-1 \rightarrow 6). In both sub-libraries, the non-randomised regions retain the wild-
20 type Zif268 framework.

The genes for each sub-library are assembled from synthetic DNA oligonucleotides by directional end-to-end ligation using short complementary DNA linkers. The oligonucleotides contain selectively randomised codons, encoding a subset of the 20 amino acids, in the appropriate positions within the zinc fingers. Assembled
25 constructs are amplified by PCR using primers containing *Not* I and *Sfi* I restriction sites, digested with the above endonucleases to produce cloning overhangs, and ligated into

similarly prepared vector Fd-Tet-SN (Choo, Y., & Klug, A. (1994) *Proc. Natl. Acad. Sci. U.S.A.* 91, 11163-11167.) Electrocompetent *E. coli* TG1 cells are transformed with the recombinant vector and plated onto TYE medium (1.5% (w/v) agar, 1% (w/v) Bactotryptone, 0.5% (w/v) Bactoyeast extract, 0.8% (w/v) NaCl) containing 15 µg/ml
5 tetracycline.

The sub-libraries are enriched for DNA-binding members by selecting against random DNA-sequences.

The 3-finger phage library is screened with 5'-biotin-(GGTTAG)₅ (Biotin-Htelo) which has been annealed in a phosphate-buffered solution containing 150 mM potassium
10 ions then immobilised on streptavidin tubes. These salt conditions are maintained throughout the selection protocol to help maintain the structural integrity of the G-quadruplex.

Phage selections are performed as follows:

Tetracycline resistant library colonies of *E. coli* TG1 cells are transferred from
15 plates into 2xTY medium (16 g/litre Bactotryptone, 10 g/litre Bactoyeast extract, 5 g/litre NaCl) containing 50 µM ZnCl₂ and 15 µg/ml tetracycline, and cultured overnight at 30°C in a shaking incubator. Cleared culture supernatant containing phage particles is obtained by centrifuging at 300 g for 5 minutes.

For the first rounds of selection, appropriate quantities of biotinylated DNA target
20 site are immobilised on streptavidin-coated tubes (Roche) in 50 µl phosphate buffer (pH 7.4) containing 50 µM ZnCl₂ and 150 mM KCl for 30 minutes at room temperature. Bacterial culture supernatant containing phage is diluted 1:10 in selection buffer (phosphate buffer pH 7.4 with 150 mM KCl) containing 50 µM ZnCl₂, 2% (w/v) fat-free dried milk (Marvel), 1% (v/v) Tween, 20 µg/ml sonicated salmon sperm DNA), and 1 ml
25 is applied to each tube. After 1 hour at 20°C, the tubes are emptied and washed 20 times

with selection buffer containing 50 μ M ZnCl₂, 2% (w/v) fat-free dried milk (Marvel) and 1% (v/v) Tween.

Retained phage are eluted in 0.1 M triethylamine and neutralised with an equal volume of 1 M Tris-HCl (pH 7.4). Logarithmic-phase *E. coli* TG1 are infected with eluted
5 phage, and cultured overnight at 30°C in 2×TY medium containing 50 μ M ZnCl₂ and 15 μ g/ml tetracycline, to amplify phage for subsequent rounds of selection.

For enrichment of the sub-libraries 1 and 2, 50 pmol of biotinylated semi-random oligonucleotides of the form 5'-

TATANNNNNNNGGCGTGTACAGTCAGCTTCAACGTC-3'

10 and 5'-TATGTGCGGNNNNNNNTCACAGTCAGTCCACACGTC-3',
respectively, are used in selection round 1. These amounts are reduced to 20 pmol and 10 pmol in rounds 2 and 3.

The heterogeneous genes from the selected clones are recovered by PCR and recombined via a *Dde*I site, present in the sequence coding for positions +4 and +5 in F2
15 of both libraries (see WO98/53057). Recombinants are then re-cloned into phage vector, as described above. Ultimately, 3×10^6 selection-enriched library members are obtained, containing randomisations over all 3 zinc fingers.

For selections against Biotin-Htelo, using the full recombined library, 100 pmol of the pre-annealed oligonucleotide is immobilised on streptavidin-coated tubes in the first
20 round. In rounds 2 and 3, selection pressure is increased by reducing the amount of target site to 50 pmol and 1 pmol, respectively. In these rounds, 50 pmol of duplex and 50 pmol single stranded competitor oligonucleotides are also added of the form:

5'-TATANNNNNNNNNNNNNTCACAGTCAGTCCACACGTC-3'. After 3 rounds of selection, *E. coli* TG1 infected with selected phage are plated. Individual
25 colonies are picked and used to prepare phage for ELISA assays and DNA sequencing.

After three rounds of selection, four different zinc finger clones are recovered and individually screened for binding to immobilised Biotin-Htelo by an ELISA assay (Choo, Y., & Klug, A. (1994) *Proc. Natl. Acad. Sci. U.S.A.* 91, 11163-11167.).

5 The four isolated clones (Gq1-4) are sequenced. The coding sequence of individual zinc finger clones is amplified by PCR from phage samples. PCR products are sequenced manually using Thermo Sequenase cycle sequencing (Amersham Life Science).

The aligned sequences are shown in Figure 3. The clones appear to have a significant degree of sequence similarity which is indicative of a successful selection process and suggests analogous functions for each clone. Control binding assays confirm
10 that neither the phage nor the Zif268 are able to bind to Biotin-Htelo.

The sequence composition of the zinc finger helices from Zif268 is also shown for comparison in Figure 3. The palindromic charge distributions of the selected zinc fingers are very different to that of Zif268. It is interesting to note that finger 2 (F2) of Gq1-4 have each selected negatively charged acidic sidechains (Asp or Glu) particularly in positions
15 labelled -1 3 and 6 (Figure 3). This pattern is unusual for DNA-binding zinc fingers as negatively charged residues are expected to repel the surface of the phosphodiester backbone. Without wishing to be bound by theory, it is possible that these acidic residues interact with guanine -NH groups which line all four grooves of an antiparallel G-quadruplex's helical core.

20 Zinc finger protein molecule(s) selected from this library bind to single stranded human telomeric DNA with an affinity comparable to that of natural transcription factors.

There is strong discrimination between the double-stranded form of the same sequence and single-stranded variants.

Example 2: Selectively Binding of G-Quadruplex DNA

The nucleic acid binding properties of zinc finger molecules produced may be analysed.

Characterisation of the binding properties of molecules Gq1-4 (see Example 1) shows that they do indeed behave very similarly. Therefore, only one phage clone (Gq1) is used to explore the binding specificity in more detail in this Example.

Phage ELISA is performed using analogues of the Biotin-Htelo oligonucleotide which contain adenine or inosine substitutions for critical guanine residues which are important for G-quadruplex formation (see Table 2). Although adenine and inosine are structurally related to guanine, both destabilise G-quadruplex formation (Williamson, J. R., Raghuraman, M. K., & Cech, T. R. (1989) *Cell* 59, 871-880.). The adenine substitution leads to a hydrogen bonding arrangement that is incompatible with G-quartet formation, while inosine lacks an N-2 exocyclic amino group required for fully stabilising such structures.

The phage ELISA used herein is adapted from previous assays (Choo, Y., & Klug, A. (1994) *Proc. Natl. Acad. Sci. U.S.A.* 91, 11163-11167.). 5'-biotinylated DNA sites are added to streptavidin-coated ELISA wells (Boehringer-Mannheim) in 50 mM potassium phosphate buffer (pH 7.5) containing 100 mM potassium chloride and 50 μ M Zinc chloride (K/Zn buffer). Phage solution [overnight bacterial culture supernatant diluted 2:10 in K/Zn buffer containing 2% (w/v) fat-free dried milk (Marvel), 1% (v/v) Tween and 20 μ g / ml sonicated salmon sperm DNA] is applied to each well (50 μ l / well). The phage are allowed to bind for 1 hour at 20°C. Unbound phage are removed by washing 6 times with K/Zn buffer containing 1% (v/v) Tween, and then 3 times with K/Zn buffer. Bound phage are detected by ELISA using horseradish peroxidase-conjugated anti-M13 IgG (Pharmacia Biotech), and the colorimetric signal is quantified using BIO KINETICS READER EL 340 (Bio-Tek Instruments).

Under the binding assay conditions used (150 mM K⁺), Gq1 has an apparent ELISA dissociation constant (K_d^E) of 26 nM for Biotin-Htelo (eg. see Table 2, Figure 7). No significant binding of Gq1 is observed for any of the guanine-substituted analogues employed, suggesting Gq1 is highly structure-specific for G-quadruplex nucleic acid.

5 A double stranded Htelo oligonucleotide ligand is made by DNA polymerase primer extension of the C-rich complementary sequence of Htelo. This complex is also analysed for binding of G1q by ELISA and exhibits no significant binding (Table 2). Therefore, although Gq1 is specific for the Htelo sequence, it cannot bind this sequence in the double-helical conformation.

10 Thus, the nucleic acid binding polypeptides bind G-quadruplex nucleic acid in a highly structure-specific manner.

The characteristics of this Example of a nucleic acid binding polypeptide are further investigated using electromobility shift assays on G-quadruplex DNAs and DMS protection of the DNA-protein complex.

15 To explore the nature of the Gq1-Htelo complex in more detail, the gene encoding Gq1 is cloned and overexpressed as a glutathione-S-transferase fusion protein ('Gq1*') in *E. coli* (Chittenden T, Livingston DM, Kaelin WG Jr (1991) *Cell* Jun 14;65(6):1073-82; Smith DB, Johnson KS (1988) *Gene* Jul 15;67(1):31-40).

20 The zinc finger gene is amplified by PCR, using 1 µl overnight bacterial culture supernatant (containing phage) as template. The primers introduced *Bam*HI sites for ligation into vector pGEX-3X (Amersham-Pharmacia). The resulting construct (Gq1*), coding for GST fused in frame with C-terminal zinc fingers, is cloned in *E. coli* TG1 and verified by DNA sequencing. Fusion protein expression is then carried out in *E. coli* BL21 DE3. Gq1* is purified from bacterial lysates by affinity chromatography using Glutathione
25 Sepharose 4 Fast Flow (Pharmacia Biotech).

The eluted protein appears as a single band of >95 % total protein on a protein gel, and corresponds to the expected molecular weight of 37 kD.

The complex between Gq1* and oligonucleotide Htelo is studied by non-denaturing gel mobility shift analysis (Cann JR (1989) *J Biol Chem Oct*

- 5 15;264(29):17032-40. Garner MM, Revzin A (1981) *Nucleic Acids Res Jul 10;9(13):3047-60*) as follows;

Gel Mobility Shift Analysis

- Binding reactions are performed in a final volume of 10 μ l, using 10 fmol of labelled oligonucleotide and various amounts of purified Gq1* in binding buffer: 20 mM
- 10 Tris-HCl pH 7.5, 1 mM EDTA, 1 mM DTT, 6% glycerol, 100 μ g / ml BSA, 1 μ g / ml calf thymus DNA, 50 μ M ZnCl₂ and KCl to 150 mM. Binding reactions are carried out at room temperature for 1 hour. The samples are loaded on a 8 % polyacrylamide (acrylamide:bisacrylamide = 33:1) non-denaturing gel. The buffer in the gel and for electrophoresis is 0.5 X TB buffer (Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) in
- 15 *Molecular Cloning; A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.). Electrophoresis is performed at 15 V/cm, for 2 hours, at 4 °C. The gels are exposed in a phosphorimager cassette and imaged (Model 425E PhosphorImager; Molecular Dynamics, Inc). The bands are quantified using Imagequant software. The fraction of DNA that is bound and free is determined after normalisation by summing the
- 20 total number of counts in each lane (Senear DF, Brenowitz M (1991) *J Biol Chem Jul 25;266(21):13661-71*). To minimise any error due to perturbation of the equilibrium under electrophoretic conditions, the fraction of free DNA is measured at various protein concentrations rather than measuring the amount of complex formed (Cann JR (1989) *J Biol Chem Oct 15;264(29):17032-40*; Garner MM, Revzin A (1981) *Nucleic Acids Res Jul*
- 25 10;9(13):3047-60). The data is plotted as \emptyset (1-fraction of free DNA) vs protein concentration to determine the K_d , which is equal to the protein concentration at which half the free DNA is bound. Equilibrium dissociation constants (K_d) are extracted by non-

linear regression using the program Origin 4.1 and the following equation (Gunasekera A, Ebright YW, Ebright RH (1992) *J Biol Chem* Jul 25;267(21):14713-20)

$$\emptyset = [P] / \{K_d + [P]\}$$

where \emptyset denotes the fractional saturation of DNA (i.e. fraction of DNA complexed with the protein). [P] represents the protein concentration in the experiment. \emptyset and [P] are inputs to the non-linear regression; K_d is an unconstrained output.

Various concentrations of Gq1* are incubated with 5' ³²-labelled-Htelo, under conditions (150 mM K⁺) that promote and stabilise the G-quadruplex conformation, and the resulting complex is run on an 8% non-denaturing polyacrylamide gel (see for example Figure 5A). This analysis shows the transition of a low molecular weight band to a single, higher molecular weight species upon increasing Gq1* concentration.

The gel mobility shift data is fitted to a quadratic (see above - Gunasekera A, Ebright YW, Ebright RH (1992) *J Biol Chem* Jul 25;267(21):14713-20) and equilibrium dissociation constants (K_d) are extracted by non-linear regression, to give an observed dissociation constant (K_d) of 34±10 nM (Figure 5B) which is close to the apparent ELISA value (K_d^E) of 26 nM. No DNA-binding is observed for GST protein alone in the absence of Gq1.

To elucidate the conformation of the oligonucleotide in the Gq1*-Htelo complex, DMS protection experiments are carried out on the complex in the form of Dimethyl sulfate protection assay of Htelo and Htelo-Gq1 zinc finger complexes.

Htelo is 5'-labelled with ³²P and is denatured by heating at 95 °C for 10 minutes. Annealing / quadruplex forming reactions are carried out as described above, in 50 mM Tris-HCl buffer with or without 150 mM potassium. DMS protection is carried out as described by Maxam and Gilbert (Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.). 1 µl of dimethylsulfate (DMS) is added to 20 pmol of annealed Htelo, at

4°C, in 200µl of appropriate buffer. The mixture is incubated at 20°C for 5 minutes. Reactions are stopped by adding 1/4 volume of stop buffer containing 1M β-mercaptoethanol and 1.5 M sodium acetate, pH 7.0. The reaction products are ethanol precipitated twice and treated with 100 µl of 1M piperidine at 90°C for 30 min. The
5 cleaved products are resolved on a 20% denaturing urea-polyacrylamide gel.

For DMS footprinting of the Htelo-Gq1 zinc finger complex, the procedure described above is adapted: 2 µl of DMS are added to 0.2 pmol of annealed Htelo, in the absence or presence of 500 nM purified Gq1* (see below), in 200 µl of the appropriate buffer, containing 1 µg/ml calf thymus DNA. Reactions are carried out for 10 minutes at
10 20°C, after which the procedure continues as described above.

Using 5' ³²-labelled-Htelo and buffer containing 100 mM K⁺, the concentration of Gq1* is set at 200 nM which is ~6-fold higher than the K_d. These conditions correspond to a near total bandshift (Figure 5A), representing complete complexation of the DNA.

In the absence of Gq1*, a cleavage protection pattern is generated that is both
15 characteristic of G-quadruplex structure, and that is dependent on the presence of 100 mM K⁺ (Figure 6; lanes 1 and 2). However, in the presence of Gq1* and 100 mM K⁺, there is still significant protection of the critical guanines (Figure 6; lane 3) indicative of G-quadruplex structure. Furthermore, in the absence of potassium, the protein does not alter the unfolded state of Htelo (Figure 6; lane 4).

20 Thus it is demonstrated that Gq1 binds Htelo in the G-quadruplex conformation, and that this nucleic acid binding polypeptide recognises the structure of folded G-quadruplex.

Example 3: Telomerase Assay

Telomerase activity may be assayed using the following method. Telomerase
25 template primers are bound to ELISA wells by biotin-streptavidin linkage as described in

Example 2. These primers are non-G-rich and are not bound by Gq1*. Test extracts are added to wells in telomerase extension buffer. The test extracts may contain telomerase activity. Such activity would cause primer extension through the addition of repeats of the sequence [(GGGTTA)_n].

- 5 A telomerase extension reaction is carried out in telomerase extension conditions. Telomerase products [(GGGTTA)_n] are detected by ELISA as described in Example 2.

This method provides a convenient and rapid technique for the assay of telomerase activity, and/or the detection of candidate telomerase activities.

10 **Example 4. Surface Plasmon Resonance Study of the Gq1-GST Binding Interaction with a G-Quadruplex: Affinity and Stoichiometry of Interaction**

The binding affinity of Gq1-GST for Htelo has previously been measured by non-denaturing gel mobility shift assay. Also, the affinity of Gq1-phage for Htelo has been measured by phage ELISA (Example 2). Both methods give similar K_d values of around 30 nM.

- 15 In this Example, another method is used to validate the binding affinity results obtained so far. In the following experiments surface plasmon resonance (SPR) is employed to determine the real time kinetics and binding affinities of the Gq1-GST-Htelo interaction. This technology allows the measurement of macromolecular interactions in real time, hence both binding and rate constants for protein binding to DNA can be
20 determined. This technique has also been used to obtain the stoichiometry of the protein-DNA interaction. The theory of SPR and the BIAcore system are explained in detail below.

Experimental Design and Data Analysis is

Biotinylated oligonucleotides are captured on a sensor chip consisting of streptavidin covalently attached to a carboxymethyl dextran hydrogel on a thin gold film. Analogues of the (GGT TAG)₅ oligonucleotide containing adenine or inosine substitutions for critical guanine residues required in G-quadruplex formation are used. As described previously, although adenine and inosine are structurally related to guanine, both destabilise G-quadruplex formation. Upon running the binding experiments, the sensograms obtained are analysed and the association rate constant (k_a), the dissociation rate constant (k_d) and the dissociation equilibrium constant (K_d) are calculated from the association and dissociation curves using the BIAevaluation software 3.0 (BIAcore). Under the binding assay conditions used (100 mM K⁺, pH 7.5), Gq1-GST is shown to have a dissociation constant for (GGT TAG) of (K_d) = 24 ± 8 nM, while the association... rate constant and the dissociation rate constant are measured as $k_a = 8.3 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ and $k_d = 0.020 \text{ s}^{-1}$ respectively (Figure 8). This gave a half life for Gq1-GST dissociation of 35 s. No significant binding of Gq1-GST is observed for any of the guanine-substituted analogues employed. Thus, the binding affinity of Gq1-GST for the analogues is greater than the detection limit of the instrument of 1 mM (BIAcore Handbook).

A double stranded oligonucleotide ligand composed of 5'-(GGT TAG)₅-3' and its complementary DNA strand is also analysed but exhibited no significant binding. This agrees with the ELISA results obtained in Chapter 2. Thus conservative changes to the parent oligonucleotide resulted in significant loss of binding to Gq1-GST. These results along with the ELISA binding assays where a similar apparent binding affinity ($K_D^E = 26 \pm 7$ nM) for Gq1-phage binding to Htelo is obtained, demonstrate that the selected zinc finger protein shows sequence specificity for the parent ligand 5'-(GGT TAG)₅-3'. The inability of Gq1-GST to bind to the control DNA sequences which had been designed with the specific aim of disrupting the G-quadruplex structure is also suggestive of G-quadruplex structure recognition by Gq1-GST.

DNA sequence (5' - 3')	k_a ($M^{-1} s^{-1}$)	k_d (s^{-1})	K_d (nM)
(GGTTAG) ₅	8.3×10^5	0.020	24
duplex	-	-	-
(GGTTAG) ₅	-	-	-
(GGTTA <u>A</u>) ₅	-	-	-
(<u>A</u> GGTTAG) ₅	-	-	-
(<u>I</u> GGTTAG) ₅	-	-	-

Table 3 Binding constants derived from the sensogram of various DNA sequences binding to Gq1-GST shown in Figure 8.

Stoichiometry of Binding

Since the response in SPR is directly related to the change in surface mass concentration of the analyte (Gq1-GST), it depends on the molecular weight of Gq1-GST in relation to the number of ligand (DNA) sites on the surface. Thus, assuming that the relationship between response and mass is the same for the ligand and the analyte (1000 RU = 1 ng/mm² for proteins) then it is possible to find the number of Gq1-GST molecules that bind to a single DNA ligand using the following equation (10) (see experimental):

$$R_{\max} = (\text{analyte MW} / \text{ligand MW}) \times (\text{ligand response}) \times (\text{valence}) \quad (10)$$

where R_{\max} is the maximum binding capacity of the surface ligand for the particular analyte in RU, *analyte MW* is the molecular weight of Gq1-GST, the *ligand MW*

is the molecular weight of the DNA, *ligand response* is the amount of DNA immobilised on the surface (in RU) and *valence* is the number of Gq1-GST molecules that can bind to a single molecule of DNA.

The data from the sensogram gave a value of 5240 RU for R_{max} . The molecular weight of Gq1-GST is 37 kDa and that of biotin-Htelo is 9730 Da. The immobilisation of the ligand corresponded to 700 RU. Substituting these values in the above equation gave the valence as 1.97. Thus, two molecules of Gq1-GST bind to one molecule of Htelo DNA. If Gq1-GST is a groove binder like all the other zinc finger proteins studied so far, it can be speculated that Gq1-GST binds in two of the four grooves present on the four faces of the G-quadruplex (chair-type G-quadruplexes have two wide and two narrow grooves).

Biological implications of Gq1-GST increasing the stability G-quadruplex DNA

Although guanine-rich regions are found in various parts of the eukaryotic genome, in order to form G-quadruplexes *in vivo* the duplex DNA needs to be at least transiently melted. Apart from the single stranded region of telomeres⁸, this might occur at the replication fork during DNA replication, when hundreds of nucleotide of single stranded DNA are exposed¹²⁵, in transcriptionally active regions, or in regions prone to local unwinding. If this does happen, and G-quadruplexes are indeed formed by the single stranded DNA, this could lead to pausing in DNA synthesis during replication and even affect transcription of a gene product. It has been speculated that these structures may provide a block during processes like transcription to stop the transcription machinery from reading the unwanted codons. Probes such as Gq1-GST may be useful for studying these processes in cell based studies and *in vitro* models.

Experimental Details and Theory of SPR

SPR studies

The SPR instrument is a BIAcore 2000 (BIAcore AB, UK) used with a SA sensor chip which consisted of thin gold film coated with a carboxymethyl dextran hydrogel. derivatised with streptavidin. Each sensor chip contained four flow cells of dimensions 2.4 x 0.5x 0.05 mm (l x w x h) with a probing spot for the SPR signal of *ca.* 0.26 mm² for each flow cell. All buffers and samples are filtered and degassed prior to use.

DNA immobilisation

The biotinylated oligonucleotides at 10 nM are injected at a flow rate of 10 ml/min across individual flowcells of two different SA sensor chips using immobilisation buffer (50 mM Tris-HCl pH 7.4, 100 mM KCl, 1 mM MgCl₂, 5 mM DTT, 50 µM ZnCl₂). This resulted in the immobilisation of 700 response units (RU) of single stranded 5'-biotin-(GGTTAG)₅, 700 RU of single stranded 5'-biotin-(AGGTTAG)₅, 700 RU of single stranded 5'-biotin-(IGGTTAG)₅, 700 RU of 5'-biotin-(GGTTAA)₅ and 2100 RU of double stranded *db Htelo*. One response unit corresponds to a surface density of DNA of approximately 1 pg/mm², thus the above levels correspond to *ca.* 70 fmol/mm². One flow cell is left underivatised to control for non-specific protein binding to the sensor chip matrix, bulk refractive index changes between the injected solution and the running buffer and baseline drift. After use, the sensor chip is washed in deionised water, dried and stored over dry silica gel at 4 °C. Reproducible levels of protein binding are maintained for at least two sets of experiments.

SPR assay

Gq1-GST is diluted two fold from 60 to 1.75 nM in running buffer (50 mM Tris-HCl pH 7.4, 100 mM KCl, 1 mM MgCl₂, 5 mM DTT, 50 µM ZnCl₂, 20 mg/ml calf thymus DNA) and injected for 240 s at a flow rate of 20 ml/min over DNA-derivatised

and control flowcells. The protein sample is then replaced by running buffer and the protein-DNA complex allowed to dissociate for 300 s. The chip surface is regenerated with an injection of 1 M NaCl for 60 s. All assays are carried out at 25 °C with data points taken every 0.5 s. The SPR data analysis is carried out using the BIAevaluation software 3.0 on the BIAcore machine.

Stoichiometry of binding

Assuming that the relationship between response and mass is the same for the ligand and the analyte ($1000 \text{ RU} = 1 \text{ ng/mm}^2$ for proteins) it is possible to find the number of Gq1 molecules that bind to a single DNA ligand using the following equation:

$$\text{Ligand sites (pmole/mm}^2\text{)} = (\text{Ligand response/Ligand MW}) \text{ Valence} \quad (14)$$

Where *valence* is the number of analyte molecules (Gq1-GST) which can bind to one ligand molecule.

$$\text{Since the Analyte response} \propto (\text{analyte MW}) (\text{analyte molecules}) \quad (15)$$

Substituting ligand sites for the analyte molecules

$$R_{\text{max}} = (\text{analyte MW/ligand MW}) (\text{ligand response}) (\text{valence}) \quad (10)$$

where R_{max} is the maximum binding capacity of the surface ligand for the particular analyte in RU.

Example 4. Inhibition of DNA Polymerase and Human Telomerase Activity by an Engineered Zinc Finger Protein that Binds G-Quadruplexes

The G-quadruplex nucleic acid structural motif is a target for designing molecules that could potentially modulate telomere length regulation, or have anti-cancer properties.

5 The engineered zinc finger protein (Gq1) binds with specificity to the intramolecular G-quadruplex formed by the human telomeric sequence 5'-(GGTTAG)5-3'. This Example demonstrates that Gq1 is able to arrest the action of a DNA polymerase, on a template containing a telomeric sequence. Inhibition occurs in a concentration-dependent manner, presumably by forming a G-quadruplex•protein complex. Furthermore,
10 Gq1 inhibits the apparent activity of the enzyme telomerase *in vitro*, with an IC₅₀ value of 74.3 ± 11.1 nM.

Using a DNA polymerase stop assay described previously (18) we study the effect that Gq1 binding has on the stability of the G-quadruplex structure. Furthermore, we use an *in vitro* assay to investigate if Gq1 can inhibit telomere synthesis by telomerase.

15 Materials and Methods

Preparation of Gq1

The glutathione S-transferase fusion of the zinc finger protein (Gq1) is purified from bacterial lysates by affinity chromatography using Glutathione Sepharose 4 Fast Flow (Pharmacia Biotech), as previously described.(21)

20 DNA Oligonucleotides

The following oligonucleotides are purchased from the Oswel DNA service (Southampton, UK): **Htemp**, 5'-(GTG CTT (GGG ATT)₄ATG ATT ATG GAC GGC TGC GA)-3'; **13-mer**, 5'-(TCG CAG CCG TCC A; **TS**, AAT CCG TCG AGC AGA

GTT)-3'; RP, 5'-(GCG CGG (CTT ACC)₃CTA ACC)-3'; ICT, 5'-(AAT CCG TCG AGC AGA GTT AAA AGG CCG AGA AGC GAT)-3'; NT, 5'-(ATC GCT TCT CGG CCT TTT)-3'; TSR8, 5'-(AAT CCG TCG AGC AGA GTT AG (GGT TAG)₇)-3'.

Annealing or Quadruplex Formation of Oligonucleotides

- 5 Oligonucleotides are diluted to 10 μ M in 50 mM Tris-HCl (pH 7.5) in the presence or absence of 100 mM KCl, as specified. Duplex annealing or quadruplex formation is carried out by heating samples to 95 °C, on a thermal heating block, and cooling to 4 °C at a rate of 2 °C/min.

Gel Mobility Shift Assay

- 10 Binding reactions are performed in a final volume of 10 μ l, using 10 fmoles of labelled oligonucleotide and varying concentrations (0 to 1 μ M) of purified Gq1 in binding buffer (50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 mM DTT, 6 % glycerol, 100 μ g/ml BSA, 1 μ g/ml calf thymus DNA, 50 μ M ZnCl₂ and 100 mM KCl). After incubating for 1 hr at room temperature, samples are loaded on a 8 % polyacrylamide
15 (acrylamide:bisacrylamide = 33:1) non-denaturing gel. 0.5 X TB is used, both in the gel and as electrophoresis buffer. Electrophoresis is performed at 15 V/cm, for 2 h, at 4 °C. The gels are exposed in a phosphorimager cassette and imaged (Model 425E PhosphorImager; Molecular Dynamics, Inc). Bands are quantified using Imagequant software. The data are plotted as \emptyset (1-fraction of free DNA) versus protein concentration
20 to determine the K_d, which is equal to the protein concentration at which half the free DNA is bound. Equilibrium dissociation constants (K_d) are extracted by non-linear regression using the program KaleidaGraph™ version 3.0.4 and the following equation:

$$\emptyset = [P] / \{K_d + [P]\}$$

- 25 where \emptyset denotes the fractional saturation of DNA (i.e. fraction of DNA complexed with the protein).(14)

Dimethyl Sulfate Protection Assay

DNA oligonucleotide **Htemp** is 5'-labelled with ^{32}P using T4 polynucleotide kinase (Sigma) and denatured by heating at 95 °C for 10 minutes. Annealing or quadruplex forming reactions are carried out as described above, in 50 mM Tris-HCl buffer (pH 7.5) in the presence or absence of 100 mM KCl. DMS protection is carried out as described by Maxam and Gilbert.(25) 1 µl of dimethylsulfate (DMS) is added to 0.2 pmoles of annealed DNA (either 'naked' or in complex with Gq1), in the presence of 1 µg/ml calf thymus DNA, at 4 °C, in 200 µl of buffer containing 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 mM DTT, 6 % glycerol, 100 µg/ml BSA, 50 µM ZnCl_2 and KCl to 100 mM. The reaction is carried out for 5 min at room temperature and stopped by adding 1/4 volume of stop buffer containing 1 M β -mercaptoethanol and 1.5 M sodium acetate, pH 7.0. The reaction products are ethanol precipitated twice and treated with 100 µl of 1 M piperidine at 90 °C for 30 min. The cleaved products are resolved on a 20 % PAGE polyacrylamide gel (8 M urea).

DNA Polymerase Stop Assay

This assay is adapted from the method described by Haiyong Han and co-workers.(18) The 13-mer primer (10 µM) is 5'-labelled with ^{32}P and mixed with the template DNA **Htemp** (10 µM) and annealed as described above. The polymerase reaction is carried out in a final volume of 20 µl, using 20 fmoles of duplex (i.e. 1 nM) and various amounts of purified Gq1 in binding buffer (50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 mM DTT, 6 % glycerol, 100 µg/ml BSA, 1 µg/ml calf thymus DNA, 50 µM ZnCl_2 and 100 mM KCl). Gq1 is incubated with the G-quadruplex of **Htemp** for 1 h at room temperature. The polymerase extension reaction is initiated by adding Klenow fragment of *E. coli* DNA polymerase I (exo⁻) (46 nM) expressed and purified as previously described,(8) dATP, dTTP, dGTP, dCTP (1 mM each) and MgCl_2 (10 mM). Reactions are incubated at room temperature for 10 min, and then stopped by adding an equal volume of stop buffer (95 % formamide, 10 mM EDTA, 10 mM NaOH, 0.1 % xylene cyanol, 0.1 %

bromophenol blue). Extension products are separated on a 20 % PAGE / 8 M urea, and gels are visualised on a phosphorimager (Molecular Dynamics).

Measurement of Telomerase Activity

Telomerase activity is determined using the TRAPEZE detection kit (Intergen Company, U.S.A.), which is a PCR based assay originally described by Kim *et al.* (17, 23) 5 The source of telomerase is S100 extracts from K562 cells (ATCC No. CCL-243) prepared as described previously. (28) The prepared cell extract is dialysed overnight at 4 °C using a 300 kDa Spectra/Por biotech cellulose ester (CE) dialysis membrane (Spectrum) to remove smaller proteins from the extract while retaining the 550 kDa 10 telomerase complex. 2 µl of the above extract is used in each assay. Various concentrations of Gq1 are pre-incubated either with or without the cell extract (in triplicate), for 10 min at ambient temperature, prior to initiating the telomerase reaction. Telomerase/Gq1 reactions are initiated by the addition of dNTP's and the TS primer as per standard protocol. Control experiments are also carried out using GST protein which had 15 been produced in the same way as Gq1 (data not shown). This control ensured that any telomerase inhibition observed is not due to any other molecule present in the purified protein sample. Reaction mixtures are incubated for 30 min at 30 °C, after which the samples are processed using a QIAquick Nucleotide Removal Kit (QIAGEN Ltd) which purifies DNA fragments by removing all the nucleotides and proteins (including Gq1) in 20 the mixture. Pure DNA is eluted with PCR-grade water and samples for the PCR reactions are prepared by the addition of *Taq* polymerase, dNTP's, TS primer, RP primer, NT primer and the ICT template as per standard protocol. The samples are transferred to a GENEAMP 2400 thermocycler (Perkin Elmer) for PCR amplification of telomerase products (two-step cycle of 30 s at 94 °C, 30 s at 59 °C for 30 cycles). Samples are 25 analysed using 8 % non-denaturing PAGE and quantitated using a Molecular Dynamics phosphorimager. The quantitation of telomerase products and the internal PCR control is as that described by Hamilton *et al.* (17) Data are normalised and plotted as telomerase activity against final Gq1 concentration. The IC₅₀ value is estimated by fitting the data to the equation $y = 100 / (1 + (x / IC_{50}))$.

Results and Discussion

To explore whether Gq1 is capable of inhibiting the copying of DNA by stabilising a telomeric G-quadruplex, a polymerase stop assay(18, 37, 39) is designed, as illustrated in Figure 9. The principle of the assay is to copy the template sequence **Htemp** that contains
5 four consecutive human telomeric repeats 5'-(TTAGGG)-3'. The 13-mer primer is annealed to the 3'-end of the template and can be extended by a DNA polymerase upon addition of the dNTPs. If complete extension of the primer occurs, a full length 50mer product is formed. However, factors that promote and stabilise intramolecular G-quadruplex formation may lead to a specific pause site on the template, resulting in the
10 formation of a truncated 23mer product. The stop site corresponds to an adenine base on **Htemp** located 3' to the first guanine base involved in G-quadruplex formation. Before investigating the potential enzyme-inhibiting properties of Gq1, it is necessary to characterise the complex formed between the zinc fingers and an oligonucleotide that could serve as a template for a polymerase stop assay (**Htemp**; see Figure 9). The
15 interaction is therefore studied by non-denaturing gel mobility shift analysis(5, 6, 10, 32) and by dimethyl sulphate (DMS) protection assays.(36)

Various concentrations of Gq1 are incubated with 5' ³²P-labelled-**Htemp** under conditions that promote and stabilise the G-quadruplex conformation (100 mM K⁺). The resulting complexes are resolved on an 8% non-denaturing polyacrylamide gel. Figure
20 10A shows that, as Gq1 protein concentration is increased, there is a decrease in the free DNA (**Htemp**) and an increase in higher molecular weight protein-DNA complexes (**Htemp**·Gq1). The mobility shift data are fitted to a hyperbolic equation(14) to give an equilibrium dissociation constant (K_d) of 30 ± 10 nM (Figure 10B), which agrees with the K_d value of 34 nM previously obtained for the binding of Gq1 to a similar sequence.(21)
25 No binding is observed for a control GST protein lacking the zinc finger fusion (data not shown).

The DNA template (**Htemp**) is expected to form a G-quadruplex secondary structure *in vitro* in the presence of 100 mM potassium ion concentration,(18) and a

dimethyl sulphate (DMS) protection assay is carried out to confirm this structure.(36) G-quadruplex formation requires Hoogsteen-type base pairing of guanines which protects N-7 of guanine against methylation, upon exposure to the potent methylating agent DMS. Quadruplexes therefore display characteristic patterns of protection against piperidine
5 cleavage of the DNA backbone at methylated guanines.(25) Figure 11 shows that the critical, quadruplex-forming, guanines of the **Htemp** template are almost completely protected from cleavage at a K^+ concentration of 100 mM (Lane 3) as compared to a Tris buffer control (Lane 4). This is consistent with the Tris buffer lacking the metal cations required to stabilise quadruplexes. By contrast, the guanines that are not involved in
10 quadruplex formation react strongly with DMS under both salt conditions. Similarly, when **Htemp** is incubated with 500 nM Gq1, in buffer containing 100 mM K^+ , there is almost complete protection of the critical guanines. Since this set of conditions corresponds to a total band shift (lane 7, Figure 10A), which reflects complete complexation of the DNA by the protein, this suggests that Gq1 is binding specifically to the G-quadruplex formed
15 within **Htemp**. These results are consistent with our previous observations reported for Gq1 binding to the human telomeric DNA sequence 5'-(GGTTAG)₅-3'.(21).

Having established that Gq1 binds to the G-quadruplex structure of **Htemp**, the polymerase stop assay is performed. The primer extension experiments are carried out with increasing concentrations of Gq1, using identical salt conditions to those in the
20 mobility shift assay (i.e. 100 mM KCl; Figure 10A). A small amount of 23-mer pause-product is observed in the absence of Gq1, indicating the position of a G-quadruplex structure in the template (Figure 12A, lane 1). There is less 50mer product and more 23-mer with increasing Gq1 concentration with almost complete pausing at 1 μ M Gq1 (Figure 12A, lane 5). The barrier to 50-mer DNA synthesis is quantitated as the ratio of
25 the band intensities of paused extension product (23mer) to the total products in the lane.(18) This ratio is plotted against the Gq1 protein concentration in the primer extension reaction (Figure 12B). The termination of DNA synthesis at the pause site increases with Gq1 concentration until the effect saturates at ~500 nM Gq1. These results are consistent with Gq1 binding and stabilising the G-quadruplex to provide a block for

polymerase extension. Similar inhibition of DNA polymerase synthesis has also been shown for small organic molecules that bind G-quadruplex DNA(18).

Telomerase Activity Assays

To explore whether Gq1 has any influence on the *in vitro* activity of human
5 telomerase, we employed the telomere repeat amplification protocol (TRAPEZE).(23) In
the standard protocol, telomerase extends an oligonucleotide template (TS primer) to form
discrete elongated telomeric products. These products are then amplified by PCR to
facilitate their detection. Due to the limitations of the PCR reaction, whereby a minimum
length of template is required for the reverse primer to hybridise and efficiently prime the
10 PCR reaction, only products that have been elongated by four or more telomeric repeats
are detected by this method. However, TRAPEZE allows a sensitive and linear response
over the range of telomerase activity used in these studies,(20) and the inclusion of an
internal amplification standard (IC) in each sample permits reproducible quantification.
Although a PCR control carried out at 1 μ M Gq1 shows that Gq1 does not directly inhibit
15 *Taq* polymerase, controls have suggested that Gq1 does inhibit the PCR amplification of
telomeric DNA (data in supplementary information). Therefore a modified TRAPeze
assay has been employed, in which proteins are removed after the telomerase/Gq1
reactions, prior to PCR detection of telomeric repeats.

In the modified assay, telomerase/Gq1 extension reactions are first carried out with
20 the exclusion of *Taq* polymerase and the PCR primers. Gq1 is subsequently removed by a
protocol that ensures the removal of proteins, salts and unincorporated dNTP's from the
reaction mixture. The purification exploits the denaturation of proteins with a high
concentration of chaotropic salts, followed by adsorption of the telomeric DNA extension
products onto a silica-gel membrane. After repeated washes to remove residual
25 contaminants and salts, the adsorbed DNA is eluted in water and a PCR reaction carried
out on the eluate to detect telomeric repeats. Using this modified protocol, telomerase
activity is evaluated in the presence of Gq1 concentrations ranging from 0 to 375 nM
(Figure 13, lanes 1-6). A control in which the telomerase extract is heat-inactivated at 90

°C for 10 min, confirmed that addition of telomeric repeats is due to enzyme activity in the extract (Figure 13, lane 7). In addition to the cell extract experiments, an eight-repeat telomeric oligonucleotide template (TSR8) is employed as a specific PCR control in the absence of telomerase (Figure 13; lanes 8 and 9). This control shows that even 2.5 μ M

5 Gq1 has a negligible effect on the PCR amplification of the 8 repeats of TSR8. The modified assay supports the conclusion that Gq1 is causing specific inhibition of telomerase-mediated extension of the TS primer. The telomerase inhibition by Gq1 is quantified as described previously,⁽¹⁷⁾ and the IC₅₀ value is calculated to be 77.1 ± 11.8 nM (Figure 14). This IC₅₀ value is higher than the measured K_d of Gq1 for Htemp ($30 \pm$

10 10 nM). This might be reflecting that a G-quadruplex structure formed during telomerase extension is less stable than the "free" G-quadruplex target used in the binding study.

Given the DNA polymerase stop-assay data, the molecular mechanism by which Gq1 inhibits extension by telomerase is likely to be through a direct interaction of Gq1 with a TS primer which has been extended by four or more telomeric repeats. This model

15 is supported by the observation that Gq1 binds the G-quadruplex form of the sequence 5'-(TTAGGG)₄-3' in Htemp with a K_D = 30 ± 10 nM (Figures 10A, 11 and 12). Gq1 could therefore bind and stabilise telomeric G-quadruplex structures in the telomerase extension reaction resulting in the formation of a trapped Gq1-G-quadruplex-telomerase complex which disallows another molecule of TS primer from being extended by telomerase.

20 Interestingly, in the telomerase assay carried out at the highest Gq1 concentration (375 nM; Figure 13, Lane 6), inhibition of telomerase extension seems to occur before four or more telomeric repeats have been added to the TS primer by telomerase. It is therefore possible that at higher levels of protein concentration, Gq1 may be binding to other telomeric secondary structures which may require less than four extended telomeric repeats

25 to form.

Conclusion

Gq1 is an artificial protein that has been engineered to bind human telomeric G-quadruplex DNA. The primer extension studies presented here, using both telomerase and

Klenow fragment of *E. coli* DNA polymerase I, suggest that Gq1 can inhibit both the synthesis and copying of telomeric DNA sequences. Since this zinc finger protein has no detectable affinity for telomeric duplex DNA, Gq1 may prove an attractive probe for carrying out cell based studies, which will form the basis for future studies.

5 Example 5. The Effect of Gq1 on Trapeze Detection Assay

The experiments described here describe the effect of Gq1 on the PCR amplification step carried out in the standard TRAPEZE assay method.

To explore whether Gq1 had any influence on the *in vitro* activity of human telomerase we first employed the telomere repeat amplification protocol (TRAPEZE; 101). In the standard protocol, telomerase extends an oligonucleotide primer (TS primer) to form elongated telomeric products. These products are then amplified by PCR to facilitate their detection. Due to the limitations of the PCR reaction whereby a minimum length of template is required for the reverse primer to hybridise to it and efficiently prime the PCR reaction, only products that have been elongated by four or more telomeric repeats are detected by this methodology. TRAPEZE allows a sensitive and linear response over the range of telomerase activity used in these studies (102), and the inclusion of an internal amplification standard (IC) in each sample permits reproducible quantification. The addition of the internal amplification standard also confirms that Gq1 does not interfere with *Taq* polymerase during amplification. A potential issue with this assay is that when examining molecules that interact directly with telomeric DNA, there exists the possibility of specific inhibition of the PCR amplification of telomeric DNA. Such an artefact would not be apparent from controls for the inhibition of *Taq* polymerase alone. We have examined the effect of Gq1 on the PCR amplification of TSR8 which has a sequence identical to the TS primer extended with eight telomeric repeats (5'-AAT CCG TCG AGC AGA GTT AG(GGT TAG)₈-3'). The results of this study are show in the gel in Figure 15 (lanes 12-18). As the concentration of Gq1 is increased from 0 to 200 nM there is a reduction in the intensity of fragments containing more than four telomeric repeats which correlates with the length of the telomere required to form an intramolecular

G-quadruplex structure which Gq1 can bind to. It is quite clear that Gq1 *does* inhibit the PCR amplification of TSR8 in a concentration-dependent manner. That the amplification of the internal control (IC) is not affected by Gq1 is clear evidence that the inhibition is also sequence specific. Given this result, the apparent inhibition of telomerase activity from TRAPeze assay could not be clearly interpreted and lead to the need for a modified assay where Gq1 is removed from the reaction mixture prior to the PCR amplification step (Figure 15 lanes 1-11). In Figure 15, Lane 1-3 show the extended telomerase product in the absence of Gq1 with the internal control marked as IC. As the concentration of Gq1 is increased up to 200 nM (Lanes 4-9) the longer telomeric extension products clearly appear to decrease in intensity. Lane 10 is a heat control to confirm that activity is due to telomerase, and lane 11 is a PCR control carried out at 1 mM Gq1 confirming that Gq1 does not inhibit with the *Taq* polymerase.

Preparation of Gq1

The glutathione S-transferase fusion of the zinc finger protein (Gq1) is purified from bacterial lysates by affinity chromatography using Glutathione Sepharose 4 Fast Flow (Pharmacia Biotech), as previously described.

DNA Oligonucleotides

The following oligonucleotides are purchased from the Oswel DNA service (Southampton, UK): TS, 5'-(AAT CCG TCG AGC AGA GTT)-3'; RP, 5'-(GCG CGG (CTT ACC)₃CTA ACC)-3'; ICT, 5'-(AAT CCG TCG AGC AGA GTT AAA AGG CCG AGA AGC GAT)-3'; NT, 5'-(ATC GCT TCT CGG CCT TTT)-3'; TSR8, 5'-(AAT CCG TCG AGC AGA GTT AG (GGT TAG)₇)-3'.

Measurement of Telomerase Activity

Telomerase activity is determined using the TRAPEZE detection kit (Intergen), which is a PCR based assay originally described by Kim *et al.* (3,32) The source of

telomerase is S100 extracts from K562 cells (ATCC No. CCL-243) prepared as described previously (104). The prepared cell extract is dialysed overnight at 4 °C using a 300 kDa Spectra/Por biotech cellulose ester (CE) dialysis membrane (Spectrum) to remove smaller proteins from the extract while retaining the 550 kDa telomerase complex. 2 µl of the
5 above extract is used in each assay. Gq1 at varying concentration is pre-incubated with or without the cell extract (in triplicate) for 10 min at ambient temperature prior to initiating the telomerase reaction by addition of dNTP's, TS primer 5'-(AAT CCG TCG AGC AGA GTT)-3', Taq polymerase, and PCR primers [PCR mix 1 containing RP + ICT + NT primers] as described in the TRAPEZE kit. Control experiments are also carried out at
10 various concentration of Gq1 where instead of telomerase, a TSR8 template containing 8 telomeric repeats is added. This control served to test if Gq1 specifically inhibits the PCR amplification of telomeric DNA.

All the above reaction mixtures are incubated for 30 min at 30 °C, after which the samples are transferred to a GENEAMP 2400 thermocycler (Perkin Elmer) for PCR
15 amplification of telomerase products (two-step cycle of 30 s at 94 °C, 30 s at 59 °C for 30 cycles). Samples are analysed using 8 % non-denaturing PAGE and quantitated using a Molecular Dynamics phosphorimager.

Example 6. *In vivo* effects of Gq1 - transfection of mammalian cells with Gq1-GFP peptide

20 In order to ascertain the *in vivo* properties of Gq1, pilot experiments are carried out in which the genes for the three fingers of Gq1 are fused to the gene for Enhanced Green Fluorescent Protein (EGFP) (Figure 16). Plasmid constructs carrying these fusions are introduced into mammalian cell lines by transient transfection, and any resulting phenotypic changes are monitored by fluorescence microscopy (Figures 17 to 20).

25 The sequence of the Gq1-NLS construct for GFP fusion is shown below. This construct contains the three zinc fingers from Gq1, an SV40 nuclear localisation signal

(PKKKRKV), XhoI and BamHI restriction sites that are in frame for cloning into the GFP fusion vector, pEGFP-N3 (Clontech Labs).

```

5      1/1                      31/11
GCg ACg GCG Gct cga gCC GCC ATG ggg ccc aag aag aag cgt aag gtc ggc ggt ggc GCg
A   T   A   A   R   A   A   M   G   P   K   K   K   R   K   V   G   G   G   A

10     61/21                      91/31
GAa GAg aGg CCc TAc GCa TGC CCT GTC GAG TCC TGC GAT CGC CGC TTT TCT gac tcg gcc
E   E   R   P   Y   A   C   P   V   E   S   C   D   R   R   F   S   D   S   A

15     121/41                      151/51
cac CTT ACC cgg CAT ATC CGC ATC CAC Acc GGt CAG AAG CCC TTC CAG TGT CGA ATC TGC
H   L   T   R   H   I   R   I   H   T   G   Q   K   P   F   Q   C   R   I   C

20     181/61                      211/71
ATG CGT AAC TTC AGT gac agg tcc gac Ctg aGc gaa CAC ATC CGC ACC CAC ACA GGC GAG
M   R   N   F   S   D   R   S   D   L   S   E   H   I   R   T   H   T   G   E

25     241/81                      271/91
AAG CCT TTT GCC TGT GAC ATT TGT GGG AGG Aaa ttt GCC cgc agc gac cac CGC ata gaa
K   P   F   A   C   D   I   C   G   R   K   F   A   R   S   D   H   R   I   E

30     301/101                      331/111
CAT ACC aag ata cac ctg cgc caa aaa gat gcg GGA TCC gag tga ttg
H   T   K   I   H   L   R   Q   K   D   A   G   S   E   *   L

```

30 The Gq1-NLS-EGFP Fusion Construct

Fusion protein constructs are made between a G-quadruplex-binding zinc finger (Gq1) and Enhanced Green Fluorescent Protein (EGFP; Clontech Labs.). A strong nuclear localisation signal (NLS) from the SV40 large T antigen, is added to complete the construct (Gq1-NLS-EGFP). As a control, a plasmid containing and expressing EGFP alone (no NLS), is also used (pEGFP-N3; Clontech Labs.).

All constructs are initially expressed in HeLa cells by transient transfection: Figure 16 shows the typical result of transfecting a HeLa cell with a control plasmid that expresses EGFP alone. Green fluorescence (indicative of EGFP) is evenly distributed in both the cytoplasmic and nuclear compartments of these cells, and there are no apparent phenotypic or morphological changes observed. By contrast, when Gq1-NLS-EGFP is transfected into HeLa cells, there are three striking differences:

(i) The green EGFP fluorescence is almost entirely localised in the nucleus, indicating the efficiency of the NLS (Figure 17).

(ii) Inside the nucleus, the zinc finger-EGFP fusions are concentrated or sublocalised within discrete regions, reminiscent of nucleoli (Figure 17).

5 (iii) The cells displayed a multilobar-nuclear or multinuclear phenotype (Figures 17, 18). This nuclear fragmentation could be due to apoptotic cell death and has been induced by the presence of Gq1-NLS-EGFP.

It is clear from these preliminary results that the Gq1-NLS-EGFP is having strong morphological and cytotoxic effects on these cells. We therefore set out to investigate the
10 mechanisms by which these effects occur.

Chromosome Staining of Cells Transfected with Gq1-NLS-EGFP

In order to define the nuclear sublocalisation of Gq1-NLS-EGFP, with respect to the chromosomal DNA, transfections are repeated with propidium iodide staining of cells, which colours DNA red. For these experiments, COS7 cells are used as they are less prone
15 to apoptosis and nuclear degradation than HeLa cells. Only thus is it possible to find cells with visible condensed chromosomes that had been transfected with pGq1-NLS-EGFP (Figure 19). By adding colcemide to the cells 24 hours after transfection, it is even possible to halt transfected cells during metaphase, the point at which paired chromosomes become aligned, immediately prior to mitotic cell division (Figure 20). From these studies,
20 it is apparent that EGFP fluorescence is not sufficiently sensitive to demonstrate conclusively that Gq1 co-localises with the telomeric ends of chromosomal DNA. Indeed, it appears that in COS7 cells the zinc finger is randomly distributed relative to chromosomal DNA. In HeLa cells, which showed a more marked phenotypic effect, no condensed chromosomes are ever seen, despite screening a large number of transfected
25 cells. These results indicate that the cytotoxic effects of Gq1-NLS-EGFP are cell-type specific.

Because of the sensitivity of the current assay, it is impossible to say whether the effects seen are mediated by interactions between the zinc fingers and telomeric DNA. In order to establish the nature of the *in vivo* interactions of Gq1 it will be necessary to develop more sensitive detection assays. For example, immunodetection of c-myc-tagged Gq1 could reliably detect the location of much smaller doses of peptide in a cell. We envisage that these ongoing studies will eventually help us to understand the *in vivo* properties of engineered zinc fingers that bind G-quadruplex DNA.

Constructing Pgq1-NLS-EGFP Fusion Plasmid

Zinc finger genes are amplified by PCR from a Gq1 phage clone, using 1 μ l overnight bacterial culture supernatant (containing phage) as template. The primers introduced *XhoI/BamHI* sites for ligation into vector pEGFP-N3 (Clontech Labs). In addition, the forward primer introduced an SV40 large T antigen nuclear localisation signal (NLS). The resulting construct (pGq1-EGFP-NLS) is cloned in *E. coli* TG1 and verified by DNA sequencing.

Transient Transfection

pGq1-EGFP-NLS and pEGFP-N3 (control) plasmids are prepared from *E. coli* TG1 using Qiagen Endotoxin-free Maxiprep Kits. All plasmids are diluted in water to 1 μ g/ μ l. Approximately 2×10^5 HeLa or COS7 cells are seeded in each well of a sterile, 6-well culture dish, containing sterile glass cover slips, and are grown for 18 hours at 37°C. For transfection, 1 μ g of each EGFP plasmid is mixed with 1 μ g of pUC-19 plasmid (carrier DNA), and with 2 μ l of lipofectamine reagent (Gibco), in a total volume of 200 μ l serum-free cell-culture medium (DMEM). This mixture is inverted vigorously several times and left to stand for 15 minutes. The mixture is then added to the cells in the 6-well dishes, already containing 800 μ l of fresh serum-free medium, and the transfection is left for 2 hours at 37°C. After this incubation, the culture medium is replaced with 3 ml of medium containing 10% fetal calf serum. Cells are allowed to grow for a further 24-48 hours, after which the glass cover slips are fixed with a mixture of paraformaldehyde (2% (v/v),

glutaraldehyde (0.2% (v/v)) in PBS. Fixing is carried out for 20 minutes at ~20°C, after which the cover slips are mounted on glass slides and examined by confocal fluorescent light microscopy.

Chromosome Staining of Cells Transfected with Gq1-NLS-EGFP

- 5 Transient transfection is carried out as described above, using COS7 cells. Cells are grown for 20 hours after transfection, and then 0.3µg of colcemide is added per 3 ml of culture medium. This metaphase block is carried out for 4 hours at 37°C, cells are then harvested by trypsinisation, and spun at 1000 rpm in a swinging-bucket centrifuge. The supernatant is entirely removed and replaced by 300µl of RSB buffer (10 mM Tris, pH
- 10 7.4; 10 mM NaCl; 5 mM MgCl₂). Cells are resuspended by gentle tapping and the mixture incubated for 10 min at 37°C.

- For chromosome spreads, a 22mm glass coverslip is placed on filter paper at the bottom of the swing out buckets of a table-top centrifuge. 50µl of cells are gently transferred onto the cover slip. The centrifuge is immediately accelerated to 3000rpm and
- 15 then stopped. Glass slides are immediately fixed in PBS with 2% (v/v) formaldehyde, for 10 minutes at ~20°C. Slides are washed once in PBS and cells are permeabilised in 0.5% NP40/PBS for 10 min. After a second wash in PBS, DNA is stained by adding 1:10 000 propidium iodide for 5 min at ~20°C. Cells are washed in PBS, mounted on glass slides and examined by confocal fluorescent light microscopy.

20 **Example 6. Gq1 as a Reagent for High-Throughput Drug Screen Assays**

- There is considerable interest in molecules that bind to telomeric DNA sequences and G-quadruplexes with specificity. Such molecules would be useful to test hypotheses for telomere length regulation, and may have therapeutic potential for diseases such as cancer. Here we describe how the Gq1 zinc finger can be employed as a tool to search for
- 25 drug-like compounds that target telomeres.

A fluorescein (FL) donor is introduced into a human telomeric DNA sequence using a fluorescein phosphoramidite during oligonucleotide synthesis:

5'- Biotin- tta ggg tta ggg tta ggg tta ggg tta ggg-FL-3'

To introduce an acceptor, there follows conjugation of Rhodamine Green TFA, succinimidyl ester to an amine (from an N-terminal lysine in the zinc finger), followed by deprotection of the fluorophore with either hydroxylamine or ammonia.

The FRET assay is essentially carried out as described by Hillisch et al., (*Curr Opin Struct Biol.* 2001 Apr 1;11(2):201-207; and Refs. therein). The assay takes place in the presence of 50-100nM of dye-labelled Gq1, in a zinc finger binding buffer (PBS containing 50 uM ZnCl₂, 1 % (v/v) Tween, 20 mg/ml sonicated salmon sperm DNA) with the addition of appropriate concentrations of the candidate small molecule. The candidate molecule is provided as part of a library; alternatively, the candidate molecule is provided in the form of an array.

Example 7. Dimers and Derivatives of Gq1

The following dimer constructs with a variety of peptide linkers are constructed

1) Construct Gq1(1:3)-linkerA-Gq1(1:3) comprising [Gq1 Fingers1-3]-linkerA - [Gq1 Fingers1-3]

2) Construct Gq1(1:3)-linkerB-Gq1(1:3) comprising [Gq1 Fingers1-3]-linkerB - [Gq1 Fingers1-3]

3) Construct Gq1(1:2)-linkerA-Gq1(1:2) comprising [Gq1 Fingers1-2]-linkerA - [Gq1 Fingers1-2]

4) Construct Gq1(1:2)-linkerB-Gq1(1:2) comprising [Gq1 Fingers1-2] - linkerB - [Gq1 Fingers1-2]

Where: linkerA = TG GGS ERP and linkerB = TG GGS GGS GGS GGS GGS ERP

5 The sequences of these constructs are shown below:

Construct Gq1(1:3)-linkerA-Gq1(1:3) nucleic acid sequence:

ATGGCGGAAGAGAGGCCCTACGCATGCCCTGTCGAGTCCTGCGATCGCCGCTTTTCTGAC
TCGGCCCACCTTACCCGGCATATCCGCATCCACACCGGTCAGAAGCCCTTCCAGTGTCGA
ATCTGCATGCGTAACTTCAGTGACAGGTCCGACCTGAGCGAACACATCCGCACCCACACA
10 GGCGAGAAGCCTTTTGCCTGTGACATTTGTGGGAGGAAATTTGCCCGCAGCGACCACCGC
ATAGAACATACCAAGATACACACAGGAGGGGGCGGATCTGAGAGGCCCTACGCATGCCCT
GTCGAGTCCTGCGATCGCCGCTTTTCTGACTCGGCCCACCTTACCCGGCATATCCGCATC
CACACCGGTCAGAAGCCCTTCCAGTGTCGAATCTGCATGCGTAACTTCAGTGACAGGTCC
GACCTGAGCGAACACATCCGCACCCACACAGGCGAGAAGCCTTTTGCCTGTGACATTTGT
15 GGGAGGAAATTTGCCCGCAGCGACCACCGCATAGAACATACCAAGATACACCTGCGCCAA
AAAGATGCGGCCGCGGAG

Construct Gq1(1:3)-linkerA-Gq1(1:3) amino acid sequence:

MAEERPYACPVESCDRRFSDSAHLTRHIRIHTGQKPFQCRICMRNFSRSDLSEHIRTHT
GEKPFACDICGRKFARSDHRIEHTKIHTGGGGSERPYACPVESCDRRFSDSAHLTRHIRI
20 HTGQKPFQCRICMRNFSRSDLSEHIRTHTGEKPFACDICGRKFARSDHRIEHTKIHLRQ
KDAAAE

Construct Gq1(1:3)-linkerB-Gq1(1:3) nucleic acid sequence:

ATGGCGGAAGAGAGGCCCTACGCATGCCCTGTCGAGTCCTGCGATCGCCGCTTTTCTGAC
TCGGCCCACCTTACCCGGCATATCCGCATCCACACCGGTCAGAAGCCCTTCCAGTGTCGA
25 ATCTGCATGCGTAACTTCAGTGACAGGTCCGACCTGAGCGAACACATCCGCACCCACACA
GGCGAGAAGCCTTTTGCCTGTGACATTTGTGGGAGGAAATTTGCCCGCAGCGACCACCGC
ATAGAACATACCAAGATACACACAGGCGGGGGCGGAAGCGGCGGAAGCGGCGGAAGCGGC
GGAAGCGGCGGATCTGAGAGGCCCTACGCATGCCCTGTCGAGTCCTGCGATCGCCGCTTT
TCTGACTCGGCCCACCTTACCCGGCATATCCGCATCCACACCGGTCAGAAGCCCTTCCAG
30 TGTCGAATCTGCATGCGTAACTTCAGTGACAGGTCCGACCTGAGCGAACACATCCGCACC
CACACAGGCGAGAAGCCTTTTGCCTGTGACATTTGTGGGAGGAAATTTGCCCGCAGCGAC
CACCGCATAGAACATACCAAGATACACCTGCGCCAAAAGATGCGGCCGCGGAG

Construct Gq1(1:3)-linkerB-Gq1(1:3) amino acid sequence:

MAERPYACPVESCDRRFSDSAHLTRHIRIHTGQKPFQCRICMRNFSDRSDLSEHIRTHT
 GEKPFACDICGRKFARSDHRIHTKIHTGGGSGGSGGSGGSGGSGSERPYACPVESCDRRF
 SDSAHLTRHIRIHTGQKPFQCRICMRNFSDRSDLSEHIRTHTGEKPFACDICGRKFARSD
 HRIHTKIHLRQKDAAAE

5 Construct Gq1(1:2)-linkerA-Gq1(1:2) nucleic acid sequence:

ATGGCGGAAGAGAGGCCCTACGCATGCCCTGTCGAGTCCTGCGATCGCCGCTTTTCTGAC
 TCGGCCCACCTTACCCGGCATATCCGCATCCACACCGGTCAGAAGCCCTTCCAGTGTCTGA
 ATCTGCATGCGTAACTTCAGTGACAGGTCCGACCTGAGCGAACACATCCGCACCCACACA
 GGAGGGGGCGGATCTGAGAGGCCCTACGCATGCCCTGTCGAGTCCTGCGATCGCCGCTTT
 10 TCTGACTCGGCCCACCTTACCCGGCATATCCGCATCCACACCGGTCAGAAGCCCTTCCAG
 TGTCTGAATCTGCATGCGTAACTTCAGTGACAGGTCCGACCTGAGCGAACACATCCGCACC
 CACCTGCGCCAAAAGATGCGGCCGCGGAG

Construct Gq1(1:2)-linkerA-Gq1(1:2) amino acid sequence:

MAERPYACPVESCDRRFSDSAHLTRHIRIHTGQKPFQCRICMRNFSDRSDLSEHIRTHT
 15 GGGGSGSERPYACPVESCDRRFSDSAHLTRHIRIHTGQKPFQCRICMRNFSDRSDLSEHIRT
 HLRQKDAAAE

Construct Gq1(1:2)-linkerB-Gq1(1:2) nucleic acid sequence:

ATGGCGGAAGAGAGGCCCTACGCATGCCCTGTCGAGTCCTGCGATCGCCGCTTTTCTGAC
 TCGGCCCACCTTACCCGGCATATCCGCATCCACACCGGTCAGAAGCCCTTCCAGTGTCTGA
 20 ATCTGCATGCGTAACTTCAGTGACAGGTCCGACCTGAGCGAACACATCCGCACCCACACA
 GCGGGGGCGGAAGCGGCGGAAGCGGCGGAAGCGGCGGAAGCGGCGGATCTGAGAGGCC
 TACGCATGCCCTGTCGAGTCCTGCGATCGCCGCTTTTCTGACTCGGCCCACCTTACCCGG
 CATATCCGCATCCACACCGGTCAGAAGCCCTTCCAGTGTCTGAATCTGCATGCGTAACTTC
 AGTGACAGGTCCGACCTGAGCGAACACATCCGCACCCACCTGCGCCAAAAGATGCGGCC
 25 GCGGAG

Construct Gq1(1:2)-linkerB-Gq1(1:2) amino acid sequence:

MAERPYACPVESCDRRFSDSAHLTRHIRIHTGQKPFQCRICMRNFSDRSDLSEHIRTHT
 GGGSGGSGGSGGSGGSGSERPYACPVESCDRRFSDSAHLTRHIRIHTGQKPFQCRICMRNF
 30 SDRSDLSEHIRT HLRQKDAAAE

The two linkers (A and B) replace the canonical zinc finger linkers (TGEKP or TGERP) to allow more flexibility of interaction. The longer linker B can span a long spatial separation between the two Gq1 molecules when bound to the DNA.

The affinity of binding between the constructs and their target G-quadruplex DNA
 35 are tested. It is found that each of these dimer constructs binds to the DNA with at least as

great an affinity as the "monomer" constructs (e.g., Gq1). Furthermore, the constructs bind DNA with sub-nanomolar affinity.

Other linkers may also be used. Such linkers may contain glycine and serine residues, preferably alternating combinations of glycine and serine residues. The linkers preferably comprise insertions of one or more glycine residues with one or more serine residues in a canonical linker. For example, a linker having the sequence TG GGGS GGGS GGGS GGGS ERP may be employed. Such a linker may also be used in place of or in addition to linker A and/or linker B.

Example 8. Use of Quadruplex Binding Polypeptides to Inhibit Viral Replication

This Example sets out assays for testing anti-HIV properties of the nucleic acid binding polypeptides described in this document, including Gq1 and its derivatives.

i. Transfection of Gq1 zinc finger DNA Constructs and Challenge with HIV-1

NP2/CD4 cells are set up at 10^5 cells per well in 6-well trays in DMEM, 5% foetal calf serum and antibiotics. NP2 cells are a human glioma cell line that do not express the common HIV and SIV coreceptors (Soda, Y., N. Shimizu, A. Jinno, H. Y. Liu, K. Kanbe, T. Kitamura, and H. Hoshino. 1999. *Establishment of a new system for determination of coreceptor usages of HIV based on the human glioma NP-2 cell line*. Biochem. Biophys. Res. Commun. 258:313-321).

The following day, various combinations of plasmid DNA are transfected with and without pCDNA3.1/Gq1 (and Gq1-derivative) expression constructs. Transfections are carried out using lipofectin (Gibco) following the maker's instructions. 1 day after transfection, the cells are trypsinised and reseeded into 48 well trays at 2.5×10^4 cells per well and reincubated.

The next day, the transfected cells are challenged with tenfold serial dilutions of the HXB2 strain of HIV-1. 100µl of virus supernatant is added to the wells and incubated for 3 hours, after which 1 ml of growth medium is added and the infected cells incubated. After 3 days, the cells are washed in PBS and fixed in cold (-40°C) methanol acetone 1:1
5 for ten minutes. After further PBS and PBS + 1% FCS washes, the cells are immunostained using p24 monoclonal antibodies, followed by an anti-mouse IgG-β-galactosidase and then enzyme substrate as described previously (Simmons, G., A. McKnight, Y. Takeuchi, H. Hoshino, and P. R. Clapham. 1995. Cell-to-cell fusion, but not virus entry in macrophages by T-cell line tropic HIV-1 strains: a V3 loop-determined
10 restriction. Virology. 209:696-700). Foci of infection stained blue and are estimated by light microscopy.

Assays are performed in duplicate, and foci of infection are counted so as to verify that the zinc finger Gq1 specifically represses HIV-1 (HXB2 strain) replication in human cell culture (Table 2 below). Repression should not occur when a control zinc finger
15 repressor (pZif268) that is specific for a different DNA sequence is used, thus showing that repression is not attributable to non-specific repression from the zinc finger domain.

ii. Delivery of Gq1 Zinc Fingers to Human Cells Using a Viral Vector

The oncoretroviral vector used contains the Gq1 gene and cis-acting viral sequences for gene expression and viral replication, such as the Long Terminal Repeat
20 (LTR), the primer binding site, the attachment site and polypurine tract sequences and an extended packaging signal. It has been deleted of all viral protein coding sequences so that it is not replication competent. This vector has been used in many gene therapy clinical trials and has shown no sign of toxicity either *ex vivo* or in patient treated.

The Gq1 gene is cloned by standard genetic engineering methods into an LNL-type
25 vector inserted into a pUC backbone. The expression of Gq1 is placed under the transcriptional control of the Moloney murine leukemia virus (Mo-MuLV) long terminal repeat (LTR). The viral vector also encodes a marker protein, the green fluorescent protein

(GFP). The expression of this marker gene is also driven by the viral LTR, a mechanism made possible by the insertion of an internal ribosomal entry site (IRES) sequence between both genes.

5 The helper functions essential to propagate the retroviral vector, such as replication and production of a functional viral capsid, may be provided by helper cells (packaging cell line) or by co-transfected plasmids.

Viral supernatant is produced by transient transfection of 293T cells, as described in detail in the following Example. The helper functions are provided from two different constructs, one expressing Gag-Pol encoding the viral capsid, reverse transcriptase and
10 integrase but lacking the encapsidation signal normally present in the Gag region and another expressing the envelope. For successful infection of human cells, the envelope used derives from the feline endogenous retrovirus (RD114) envelope protein but alternatively the Gibbon Ape Leukemia virus (GALV) envelope protein or the G protein of vesicular stomatitis virus (VSV-G) may be used.

15 Oncoretroviral Vector Production

RD114 pseudotyped vectors are produced by transient transfection of three plasmids into 293T cells: the transfer vector plasmid (LNL-based), pHIT60 (from Prof Mary Collins' lab, UCL, London, UK) a helper packaging plasmid encoding GAG and
POL proteins of murine leukemia virus, and pRDF (from Prof Mary Collins' lab, UCL,
20 London, UK) encoding for feline endogenous retrovirus (RD114) envelope protein.

A total of 1.5×10^7 293T cells are seeded in one 150-cm² flask over-night prior to transfection. Cells are cultured at 37°C in Dulbecco's modified Eagle medium (DMEM) with 10% fetal calf serum (FCS) in a 5% CO₂ incubator. A total of 72 µg of plasmid DNA is used for the transfection of one flask: 12 µg of the envelope plasmid (pRDF), 24 µg of
25 packaging plasmid (pHIT60), and 36 µg of transfer vector (pRetro) plasmid are pre-complex with lipofectamine 2000 (life technology) in Optimem according to the

manufacturer instructions. The DNA plus lipofectamine complexes are then added to the cells. After 4 hours incubation at 37 °C in a 5% CO₂ incubator, the medium is replaced by fresh DMEM or alternatively RPMI supplemented with 10% FCS and further incubated at 33°C to enhance the stability of the recombinant virus. At 36 hours and 60 hours post-
5 transfection, the medium is harvested, cleared by low-speed centrifugation (1200 rpm, 5 min), filtered through 0.45-µm-pore-size filters and use directly or kept at -80 °C.

Transduction of Human Cells

Hela and Jurkat cell are then infected with the recombinant viral vector encoding the Gq1 gene. An empty viral vector containing the GFP gene is used as control.

10 Hela cell line, a human cell line, is grown according to supplier instruction in DMEM L-glutamine containing medium supplemented with penicillin/streptavidin and fetal calf serum (complete DMEM). For successful infection with the recombinant viral vector, cells are harvested using trypsin /EDTA and 10⁵ cells are plated into a 6 well-cell culture plate containing 4 ml of viral supernatant. Cells are then further incubated for three
15 to five days at 33°C in 5% CO₂.

20 The Jurkat T cell line, a human derived lymphoblast T cell, is grown according to supplier instruction in RPMI 16100 L-glutamine containing medium supplemented with penicillin/streptavidin and fetal calf serum (complete RPMI). Cells are resuspended in 3 ml of freshly harvested retroviral supernatant and added at the concentration of 10⁵/well to a 6 well non-tissue culture treated plate (Becton Dickinson) pre-coated with 15µg/cm² retronectin (TaKaRa, Shiga, Japan). Plates are then incubated for 16 hours at 33°C. A total of 2 rounds of infection are performed in which two-third of the medium is replaced with viral supernatant. At the end of the transduction protocol cells are harvested using complete RPMI.

iii. Detection of Gq1 Protein in Transduced Cells

After three to five days post infection, the successful delivery of the Gq1 construct into Hela and Jurkat T-cells is assayed by immunochemistry (Figure 17).

HeLa cells, used as control, are transfected by electroporation with 20µg pcmv-Gq1. These cells are seeded along with viral infected HeLa cells expressing Gq1, control viral infected HeLa cells not expressing Gq1 and Uninfected HeLa cells, at 2.5×10^5 cells per well into 2 wells each of an 8-well chamber slide (Life Technologies). The cells are incubated at 37°C, 5% CO₂ for 16 hrs.

Media is removed from each well and the cells washed twice per well with phosphate buffered saline (PBS). Samples are fixed for 20 minutes at 4°C in 4% paraformaldehyde in PBS then washed twice with PBS. Samples are permeabilised for 10 minutes at 22°C in 0.25% triton-X100 in PBS and washed twice with PBS. Samples are blocked for 15 minutes at 22°C in 10% foetal calf serum (FCS) in PBS, then incubated with mouse monoclonal anti-c-Myc antibody (Autogen bioclear UK Ltd, Wiltshire), diluted according to the manufacturers' instructions in 10% FCS in PBS, for 90 minutes at 4°C. Samples are washed with PBS then incubated with Texas Red labelled anti-mouse IgG antibody (Vector Laboratories, CA), diluted according to the manufacturers' instructions in 10% FCS in PBS, for 60 minutes at 4°C. The cells are washed for a final time in PBS, then wells and gaskets removed. Samples are dried at 22°C, mounted under a coverslip using vectashield mounting medium (Vector Laboratories, CA) and analysed under a fluorescent microscope.

iv. Protocol for Transduction of Peripheral Blood CD4⁺ T Lymphocytes (Gene Therapy)

Peripheral blood mononuclear cells (PBMCs) from each patient are selected by standard procedure. PBMCs (approximately 10^8 mononuclear/kg) are taken from the patient by leukapheresis to obtain sufficient cells for infusion. This apheresis product is

overlaid onto a Ficoll-Hypaque density gradient and centrifuged to remove any erythrocytes and neutrophils. The harvested PBMCs are depleted of CD8⁺ lymphocytes using for example an anti-CD8⁺ antibody-coated AIS MicroCel-lector™ flasks, thereby leaving a CD4⁺ enriched cell population which will be stimulated with OKT3 (anti-CD3) antibody.

Activated CD4⁺ T cell are grown and transduced in close systems such as the "Peripheral Blood Lymphocyte-MPS" (cellco Cell Max™ artificial capillary system) or alternatively in the gas permeable Lifecell® X-fold™ bags (Nexell Therapeutics Inc) pre-coated with retronectin™ (TaKaRa, Shiga, Japan). For transduction, cells are exposed to GMP-grade viral conditioned medium containing IL-2 (100U/ml) once or twice a day for two or three consecutive days. At the end of the transduction protocol, cells are harvested and re-infused into the patients (up to 10⁶ CD4⁺ T cells/kg).

v. Protocol for Transduction of CD34⁺ Repopulating Cells (Gene Therapy)

CD34⁺ repopulating cells are selected and transduced according to standard protocols. Marrow CD34⁺ or alternatively mobilised peripheral CD34⁺ cells are positively selected by an immunomagnetic procedure (CliniMACS, Miltenyi Biotec, Bergish Gladbach, Germany). CD34⁺ enriched cells are cultured in gas-permeable stem cell culture containers Lifecell® X-fold™ bags (Nexell Therapeutics Inc) pre-coated with retronectin™ (TaKaRa, Shiga, Japan) in serum free medium (X-VIVO 10 or CellGro, Biowhittaker Walkerville, MD) supplemented with cytokines such as stem cell factor (Amgen), IL-3 (Novartis), IL-6 (R&D Systems) and Flt3-L (R&D Systems). For transduction, cells are exposed to GMP-grade viral conditioned medium containing cytokines once or twice a day up to two consecutive days following the activation period. At the end of the transduction protocol, cells are harvested and infused into the patients (approximately 2-4 10⁷ cells/kg).

vi. General Protocol for HIV Infection of Transduced Cells

To determine whether cells transduced with zinc finger (Gq1 and derivative) repressor constructs are restricted with respect to the expression of HIV, cells are infected with the virus and expression of HIV is assayed via expression of p24 viral antigen as well
5 as cell viability.

Jurkat cells transduced with various retroviral vectors and expressing different zinc fingers (3 positive and one negative) or untransduced Jurkat cells are infected with HIV-1 (strains RF, HXB2 or MN) at four different multiplicities of infection (10-fold dilution series). After virus absorption for 2 hours at room temperature, the cells are washed three
10 times and distributed into duplicate wells of a 48 well cell culture plate (1×10^5 cells per well in 1ml of culture fluid). 200 μ l of culture fluid is removed from each well and replaced with 200 μ l of fresh medium daily, from day 3 until day 7. The harvested culture fluid is then assayed at different dilutions to quantitate levels of p24 viral antigen using a commercial ELISA (Abbott). In addition and in parallel, cells are distributed into duplicate
15 wells of a 96 well plate (5×10^4 cells per well in 200 μ l of medium) and incubated for 6 days prior to the addition of XTT to determine cell viability.

For each virus which is tested, the Virus Input (TCID₅₀) is assayed at the various different dilutions of no virus, 1:100, 1:1000, 1:10000 and 1:100000 for each of the following combinations: Jurkat, Jurkat + vector A, Jurkat + vector B Jurkat + vector C and
20 Jurkat + negative vector.

vii. Inhibition of HIV-1 Replication in Human T-Cells with a Stable Integrated Gq1 Zinc Finger Repressor

Human Jurkat T-cells cultured in RPMI with 10% FCS are transduced with LNL-derived retrovirus that expresses the zinc finger repressor protein pGq1 (or derivative; see
25 above Example ii. "Delivery of Zinc Fingers to Human Cells Using a Viral Vector"). Seven days after transduction, the infected cells are sorted for expression of the Gq1 zinc

finger and a pool of the cells expressing the zinc finger is made, JurkatGq1. This population is assayed by FACS analysis to verify expression of CD4/CXCR4 coreceptors against a control Jurkat cell line.

JurkatGq1 and a control Jurkat cell line are seeded into 48 well plates at 2.5×10^4 cells/well and infected with tenfold serial dilutions of the HXB2 strain of HIV-1. 100 μ l of virus supernatant is added to the wells and incubated for 3 hours followed by three washes with 1 ml of growth media. 1 ml of growth media is finally added to the cells and the cells are incubated. Daily measurements of soluble p24 antigen are made by ELISA from the culture supernatants for up to seven days. Comparison of the p24 antigen levels between the control and test cell lines can measure the inhibition of HIV-1 replication in human T-cells.

FURTHER ASPECTS OF THE INVENTION

Further aspects of the invention are now set out in the following numbered paragraphs; it is to be understood that the invention encompasses these aspects:

Paragraph 1. An isolated or purified molecule capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid.

Paragraph 2. A molecule according to Paragraph 1 wherein said nucleic acid is not in a double-helical conformation.

Paragraph 3. A molecule according to Paragraph 1 wherein said nucleic acid comprises single-stranded DNA.

Paragraph 4. A molecule according to Paragraph 1 wherein said nucleic acid is comprised in a chromosome end.

Paragraph 5. A molecule according to Paragraph 1 wherein said nucleic acid is comprised in a telomeric structure.

Paragraph 6. A molecule according to Paragraph 1 wherein said nucleic acid is in a non-Watson-Crick base paired conformation.

5 Paragraph 7. A molecule according to Paragraph 1 wherein said nucleic acid comprises Hoogsteen base pairing.

Paragraph 8. A molecule according to claim 1 wherein said molecule is a polypeptide.

10 Paragraph 9. A molecule according to Paragraph 1 wherein said molecule is a polypeptide comprising at least one zinc finger motif.

Paragraph 10. A molecule according to Paragraph 1 wherein said molecule has an affinity for G-quadruplex nucleic acid which is different from its affinity for duplex nucleic acid.

15 Paragraph 11. A method for assaying telomerase activity, said method comprising:
(i) providing a sample of nucleic acid substrate for telomerase; (ii) contacting said nucleic acid sample with a telomerase; (iii) contacting said nucleic acid sample with a molecule according to Paragraph 1; and (iv) monitoring the binding of said molecule to said telomerase treated nucleic acid sample.

20 Paragraph 12. A method according to Paragraph 11 wherein said assay method comprises an ELISA assay.

Paragraph 13. A method according to Paragraph 11 wherein said assay method is in micro-well format.

Paragraph 14. A method for estimating the length of telomere(s), said method comprising: (i) contacting said telomere(s) with a molecule according to Paragraph 1; (ii) monitoring the binding of said molecule to said telomere sample, and (iii) estimating the length of said telomeres from the strength of said binding.

5 Paragraph 15. A method according to Paragraph 14 wherein said method comprises an ELISA assay.

Paragraph 16. A method according to Paragraph 14 wherein said method is in micro-well format.

10 Paragraph 17. A method for discriminating between duplex and quadruplex nucleic acid comprising contacting a sample of nucleic acid with a molecule according to Paragraph 10, and monitoring the binding of said molecule to said nucleic acid.

Paragraph 18. A method according to Paragraph 17 wherein said method comprises an ELISA assay.

15 Paragraph 19. A method according to Paragraph 17 wherein said method is in micro-well format.

Paragraph 20. A method for detecting telomeric structures *in vivo* comprising (i) contacting a labelled molecule according to any preceding Paragraph with a sample, and (ii) monitoring said labelled molecule.

20 Paragraph 21. A method according to Paragraph 20 wherein said method comprises an ELISA assay.

Paragraph 22. A method according to Paragraph 20 wherein said method is in micro-well format.

Paragraph 23. A method for manipulating telomeric structure(s) *in vivo* comprising contacting a labelled molecule according to any preceding Paragraph with a telomeric structure, wherein said molecule further comprises an effector domain.

REFERENCES

- 5 (1) Zakian, V. A. *Science* 1995, 270, 1601.
- (2) Lingner, J.; Cooper, J. P.; Cech, T. R. *Science* 1995, 269, 1533-1534.
- (3) Kim, N. W.; Piatyszek, M. A.; Prowse, K. R.; Harley, C. B.; West, M. D.; Ho, P. L. C.; Coviello, G. M.; Wright, W. E.; Weinrich, S. L.; Shay, J. W. *Science* 1994, 266, 2011-2015.
- 10 (4) Hahn, W. C.; Counter, C. M.; Lundberg, A. S.; Beijersbergen, R. L.; Brooks, M. W.; Weinberg, R. A. *Nature* 1999, 400, 464-468.
- (5) Greider, C. W.; Blackburn, E. H. *Cell* 1985, 43, 405-413.
- (6) Bodnar, A. G.; Ouellette, M.; Frolkis, M.; Holt, S. E.; Chiu, C. P.; Morin, G. B.; Harley, C. B.; Lichtsteiner, S.; Wright, W. E. *Science* 1998, 279, 349-352.
- 15 (7) Hamilton, S. E.; Simmons, C. G.; Kathiriya, I. S.; Corey, D. R. *Chemistry and Biology* 1999, 6, 343-351.
- (8) Hahn, W. C.; Stewart, S. A.; Brooks, M. W.; York, S. G.; Eaton, E.; Kurachi, A.; Beijersbergen, R. L.; Knoll, J. H. M.; Meyerson, M.; Weinberg, R. A. *Nature Medicine* 1999, 5, 1164-1170.
- 20 (9) Shamas, M. A.; Simmons, C.; Corey, D. R.; Reis, R. J. S. *Oncogene* 1999, 18, 6191-6200.

- (10) Blackburn, E. H. *Nature* 1991, 350, 569-572.
- (11) Henderson, E. R.; Blackburn, E. H. *Mol. Cel. Biol.* 1989, 9, 345-348.
- (12) Shippenlantz, D.; Blackburn, E. H. *Science* 1990, 247, 546-552.
- (13) Greider, C. W. *Mol. Cel. Biol.* 1991, 11, 4572-4580.
- 5 (14) Blackburn, E. H. *J. Biol. Chem.* 1990, 265, 5919-5921.
- (15) Zahler, A. M.; Williamson, J. R.; Cech, T. R.; Prescott, D. M. *Nature* 1991, 350, 718-720.
- (16) Salazar, M.; Thompson, B. D.; Kerwin, S. M.; Hurley, L. H. *Biochemistry* 1996, 35, 16110-16115.
- 10 (17) Fletcher, T. M.; Sun, D. K.; Salazar, M.; Hurley, L. H. *Biochemistry* 1998, 37, 5536-5541.
- (18) Griffith, J. D.; Comeau, L.; Rosenfield, S.; Stansel, R. M.; Bianchi, A.; Moss, H.; deLange, T. *Cell* 1999, 97, 503-514.
- (19) Sundquist, W. I.; Klug, A. *Nature* 1989, 342, 825-829.
- 15 (20) Williamson, J. R.; Raghuraman, M. K.; Cech, T. R. *Cell* 1989, 59, 871-880.
- (21) Sen, D.; Gilbert, W. *Nature* 1988, 334, 364-366.
- (22) Balagurumoorthy, P.; Brahmachari, S. K. *J. Biol. Chem.* 1994, 269, 21858-21869.

- (23) Mergny, J. L.; Mailliet, P.; Lavelle, F.; Riou, J. F.; Laoui, A.; Helene, C. *Anti-Cancer Drug Design* 1999, 14, 327-339.
- (24) Kerwin, S. M. *Curr. Pharm. Design* 2000, 6, 441-471.
- (25) Ren, J.; Chaires, J. B. *Biochemistry* 1999, 38, 16067-16075.
- 5 (27) Han, H. Y.; Hurley, L. H.; Salazar, M. *Nucleic Acids Res.* 1999, 27, 537-542.
- (28) Gunasekera, A.; Ebright, Y. W.; Ebright, R. H. *J. Biol. Chem.* 1992, 267, 14713-14720.
- (29) Maxam, A. M.; Gilbert, W. *Methods Enzymol.* 1980, 65, 499-560.
- (30) Derbyshire, V.; Freemont, P. S.; Sanderson, M. R.; Beese, L.; Friedman, J. M.;
10 Joyce, C. M.; Steitz, T. A. *Science* 1988, 240, 199-201.
- (31) Hamilton, S. E.; Pitts, A. E.; Katipally, R. R.; Jia, X. Y.; Rutter, J. P.; Davies, B. A.; Shay, J. W.; Wright, W. E.; Corey, D. R. *Biochemistry* 1997, 36, 11873-11880.
- (32) Prowse, K. R.; Avilion, A. A.; Greider, C. W. *Proc. Natl. Acad. Sci. U.S.A.* 1993, 90, 1493-1497.
- 15 (33) Holt, S. E.; Shay, J. W.; Wright, W. E. *Nature Biotechnol.* 1996, 14, 1-4.
101. Kim, N. W.; Piatyszek, M. A.; Prowse, K. R.; Harley, C. B.; West, M. D.; Ho, P. L. C.; Coviello, G. M.; Wright, W. E.; Weinrich, S. L.; Shay, J. W. *Science* 1994, 266, 2011-2015.
102. Holt, S. E.; Shay, J. W.; Wright, W. E. *Nature Biotechnol.* 1996, 14, 1-4.

103. Prowse, K. R.; Avilion, A. A.; Greider, C. W. *Proc. Natl. Acad. Sci. U.S.A.* 1993, 90, 1493-1497.

SEQUENCE LIST

SEQ. ID. No. 1
 5 GGTTAG GGTTAG GGTTAG GGTTAG GGTTAG
 SEQ. ID. No. 2
 TATANNNNNNNGGCGTGTACAGTCAGCTTCAACGTC
 SEQ. ID. No. 3
 TATGTGCGGNNNNNNNTCACAGTCAGTCCACACGTC
 10 SEQ. ID. No. 4
 TATANNNNNNNNNNNNNTCACAGTCAGTCCACACGTC

Each of the applications and patents mentioned above, and each document cited or referenced in each of the foregoing applications and patents, including during the prosecution of each of the foregoing applications and patents ("application cited
 15 documents") and any manufacturer's instructions or catalogues for any products cited or mentioned in each of the foregoing applications and patents and in any of the application cited documents, are hereby incorporated herein by reference. Furthermore, all documents cited in this text, and all documents cited or referenced in documents cited in this text, and any manufacturer's instructions or catalogues for any products cited or mentioned in this
 20 text, are hereby incorporated herein by reference. In particular, we hereby incorporate by reference International Patent Application Numbers PCT/GB00/02080, PCT/GB00/02071, PCT/GB00/03765, United Kingdom Patent Application Numbers GB0001582.6, GB0001578.4, and GB9912635.1 as well as US09/478513.

Various modifications and variations of the described methods and system of the
 25 invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in
 30 molecular biology or related fields are intended to be within the scope of the following claims.

CLAIMS

1. Use of a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid as an inhibitor of enzymatic activity.
2. A method of inhibiting an enzymatic activity, the method comprising: (a) providing an enzyme; and (b) contacting the enzyme with a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid.
3. A use according to Claim 1 or a method according to Claim 2, which comprises the step of providing a telomeric, G-quadruplex, or G-quartet nucleic acid and contacting the nucleic acid with the enzyme and/or the nucleic acid binding polypeptide.
4. A use or method according to any preceding claim, in which the enzymatic activity is selected from the group consisting of: a telomerase activity, a polymerase activity, an integrase activity and a gp120 activity.
5. A use or method according to any preceding claim, in which the enzymatic activity is inhibited *in vivo*.
6. A method of preventing replication of a retrovirus, the method comprising exposing the retrovirus or a nucleic acid portion thereof to a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid.
7. A method of treatment of a patient suffering from a disease, the method comprising administering to a patient in need of such treatment a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid.

8. A method according to Claim 6 or 7, in which the retrovirus is Human Immunodeficiency Virus, or in which the disease comprises infection by Human Immunodeficiency Virus infection.
9. A method according to Claim 7, in which the disease comprises a
5 hyperproliferative disease, preferably cancer.
10. A method for assaying telomerase activity, the method comprising:
 - (i) providing a nucleic acid substrate for telomerase;
 - (ii) contacting the nucleic acid substrate with a telomerase;
 - (iii) contacting the nucleic acid substrate with a nucleic acid binding polypeptide
10 capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid; and
 - (iv) monitoring the binding of the nucleic acid binding polypeptide to the nucleic acid substrate.
11. A method for determining the length of a telomere, the method comprising:
 - (i) contacting the telomere with a nucleic acid binding polypeptide capable of
15 binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid;
 - (ii) monitoring the binding of the nucleic acid binding polypeptide to the telomere, and
 - (iii) determining the length of the telomeres from the strength of the binding.
- 20 12. A method for discriminating between duplex and quadruplex nucleic acid comprising contacting a sample of nucleic acid with a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid, and monitoring the binding of the nucleic acid binding polypeptide to the nucleic acid.

13. A method of detecting telomeric structures in a system, the method comprising:
- (a) exposing the system to a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid;
 - (b) detecting binding between the nucleic acid binding polypeptide and any telomeric structures in the system.
14. A method according to Claim 13, in which the nucleic acid binding polypeptide is labelled.
15. A method according to Claim 13 or 14, in which the location of binding is detected to localise telomeric structures in the system.
16. A method according to Claim 13, 14 or 15, in which the system comprises a cell and binding is detected *in vivo* or *in situ*.
17. A method of identifying a molecule capable of binding to a telomeric, G-quadruplex, or G-quartet structure in a nucleic acid, the method comprising:
- (a) providing a nucleic acid comprising a telomeric, G-quadruplex, or G-quartet structure;
 - (b) providing a nucleic acid binding polypeptide capable of binding to a nucleic acid comprising such a structure;
 - (c) contacting either or both of the nucleic acid and the nucleic acid binding polypeptide with a candidate molecule; and
 - (d) determining the binding between the nucleic acid and the nucleic acid binding polypeptide.

18. A method of identifying a molecule capable of binding to a telomeric, G-quadruplex, or G-quartet structure in a nucleic acid, the method comprising monitoring the binding between a nucleic acid comprising a telomeric, G-quadruplex, or G-quartet structure and a nucleic acid binding polypeptide capable of binding to a nucleic acid comprising such a structure, in the presence and absence of a candidate molecule.

19. A method of identifying a molecule capable of binding to a telomeric, G-quadruplex, or G-quartet structure in a nucleic acid, the method comprising providing a complex between a nucleic acid comprising a telomeric, G-quadruplex, or G-quartet structure and a nucleic acid binding polypeptide capable of binding to a nucleic acid comprising such a structure; contacting either or both members of the complex with a candidate molecule; and detecting a dissociation between the members of the complex.

20. A method according to any of Claims 17 to 19, in which the candidate molecule is provided in the form of a library of candidate molecules, preferably an array of candidate molecules.

21. A method according to any of Claims 17 to 20, which further comprises a step of isolating, synthesising and/or providing a composition comprising the candidate molecule identified to have such activity.

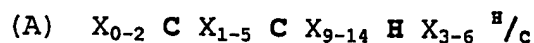
22. A method according to any of Claims 10 to 21, in which binding of the nucleic acid binding polypeptide to the nucleic acid, or the dissociation between the two, is monitored by an ELISA assay.

23. A method according to any of Claims 10 to 22, in which the binding or dissociation is monitored by detecting Fluorescence Resonance Energy Transfer (FRET).

24. A method according to any of Claims 10 to 24, in which the binding or dissociation is monitored in a micro-well.

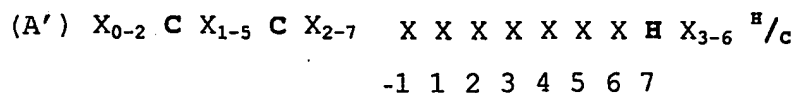
25. A method for manipulating telomeric structure(s) *in vivo* comprising contacting a labelled nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid with a telomeric structure, in which the nucleic acid binding polypeptide further comprises an effector domain.
- 5 26. A nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid for use in a method of treatment of a disease.
27. Use of a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid for the preparation of a pharmaceutical composition for the treatment of a disease.
- 10 28. A nucleic acid binding polypeptide according to Claim 26 or a use according to Claim 27, in which the disease comprises a retroviral infection, infection with Human Immunodeficiency Virus, AIDS, cancer or a hyperproliferative disease.
- 15 29. A use, method or a nucleic acid binding polypeptide according to any preceding claim, in which the nucleic acid is not in a double-helical conformation, or in which the nucleic acid binding polypeptide is capable of binding to such a nucleic acid.
30. A use, method or a nucleic acid binding polypeptide according to any preceding claim, in which the nucleic acid comprises single-stranded DNA, or in which the nucleic acid binding polypeptide is capable of binding to such a nucleic acid.
- 20 31. A use, method or a nucleic acid binding polypeptide according to any preceding claim, in which the nucleic acid is comprised in a chromosome end, or in which the nucleic acid binding polypeptide is capable of binding to such a nucleic acid.

32. A use, method or a nucleic acid binding polypeptide according to any preceding claim, in which the nucleic acid is comprised in a telomeric structure, or in which the nucleic acid binding polypeptide is capable of binding to such a nucleic acid.
33. A use, method or a nucleic acid binding polypeptide according to any preceding claim, in which the nucleic acid is in a non-Watson-Crick base paired conformation, or in which the nucleic acid binding polypeptide is capable of binding to such a nucleic acid.
34. A use, method or a nucleic acid binding polypeptide according to any preceding claim, in which the nucleic acid comprises Hoogsteen base pairing, or in which the nucleic acid binding polypeptide is capable of binding to such a nucleic acid.
35. A use, method or a nucleic acid binding polypeptide according to any preceding claim, in which the nucleic acid binding polypeptide has an affinity for G-quadruplex nucleic acid which is different from its affinity for duplex nucleic acid.
36. A use, method or a nucleic acid binding polypeptide according to any preceding claim, in which the nucleic acid binding polypeptide comprises a zinc finger motif.
37. A use, method or a nucleic acid binding polypeptide according to any preceding claim, in which the nucleic acid binding polypeptide comprises any of the following structures:

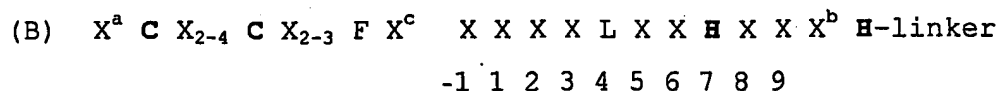


where X is any amino acid, and the numbers in subscript indicate the possible numbers of residues represented by X;

129



where X is any amino acid, and the numbers in subscript indicate the possible numbers of residues represented by X; or



where X (including X^a , X^b and X^c) is any amino acid. X_{2-4} and X_{2-3} refer to the presence of 2 or 4, or 2 or 3, amino acids, respectively.

- 5 38. A use, method or a nucleic acid binding polypeptide according to Claim 37, in which the amino acids at positions -1, 1, 2, 3, 4, 5 and 6 are selected from the group consisting of: RDSAHLTR, DRSDLSE, RSDHRIE, RSDHLIN, DRADLSE, TSSHRTN, DSAHLTR, DRDHLSE, TSSHRTN, TSHHLIQ, DRADLSE, and HQHYRTN.
- 10 39. A use, method or a nucleic acid binding polypeptide according to Claim 37 or 38, in which the polypeptide comprises three zinc finger motifs F1, F2 and F3, in which the amino acids at positions -1, 1, 2, 3, 4, 5 and 6 of F1, F2 and F3 comprise: F1: DSAHLTR, F2: DRSDLSE, F3: RSDHRIE.
- 15 40. A use, method or a nucleic acid binding polypeptide according to any preceding claim, in which the nucleic acid binding polypeptide comprises a sequence derived from at least one of the fingers of Gq1.
41. Use of a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid as a cytotoxic agent.
- 20 42. A method of killing a cell, preferably by induction of apoptosis in the cell, which method comprises exposing a cell to a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid.

43. A nucleic acid binding polypeptide comprising a sequence selected from the group consisting of: Gq1(1:3)-linkerA-Gq1(1:3) amino acid sequence, Gq1(1:3)-linkerB-Gq1(1:3) amino acid sequence, Gq1(1:2)-linkerA-Gq1(1:2) amino acid sequence, Gq1(1:2)-linkerB-Gq1(1:2) amino acid sequence, and fragments or derivatives of the
5 above.

44. A nucleic acid sequence capable of encoding a nucleic acid binding polypeptide according to Claim 42.

45. A nucleic acid sequence according to Claim 44, which is selected from the group consisting of: Gq1(1:3)-linkerA-Gq1(1:3) nucleic acid sequence, Gq1(1:3)-linkerB-Gq1(1:3) nucleic acid sequence, Gq1(1:2)-linkerA-Gq1(1:2) nucleic acid sequence, Gq1(1:2)-linkerB-Gq1(1:2) nucleic acid sequence, and fragments or derivatives of the
10 above.

46. A use, method or a nucleic acid binding polypeptide according to any preceding claim, in which the nucleic acid binding polypeptide comprises a polypeptide according to
15 Claim 43, or a polypeptide encoded by a nucleic acid sequence according to Claim 44 or 45.

1 / 15

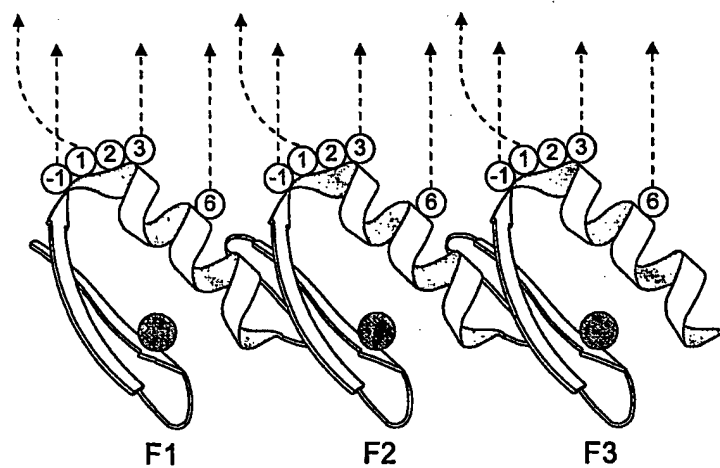


FIG. 1A

F1							F2							F3						
-1	1	2	3	4	5	6	-1	1	2	3	4	5	6	-1	1	2	3	4	5	6
A	S	A	A	L	I	A	A	S	A	A	L	S	A	A	K	A	A	R	I	A
D		D	D		T	E	D	R	D	D			E	D	N	D	D		K	E
H		H	H			K	H		H	H			K	H	R	H	H		T	K
N		K	N			N	N		K	N			N	N	S	K	N			N
Q		N	S			Q	Q		N	S			Q	Q		N	S			Q
R		Q	T			R	R		Q	T			R	R		Q	T			R
T		R	V			T	T		R	V			T	T		R	V			T
		S				V			S				V			S				V

FIG. 1B

2 / 15

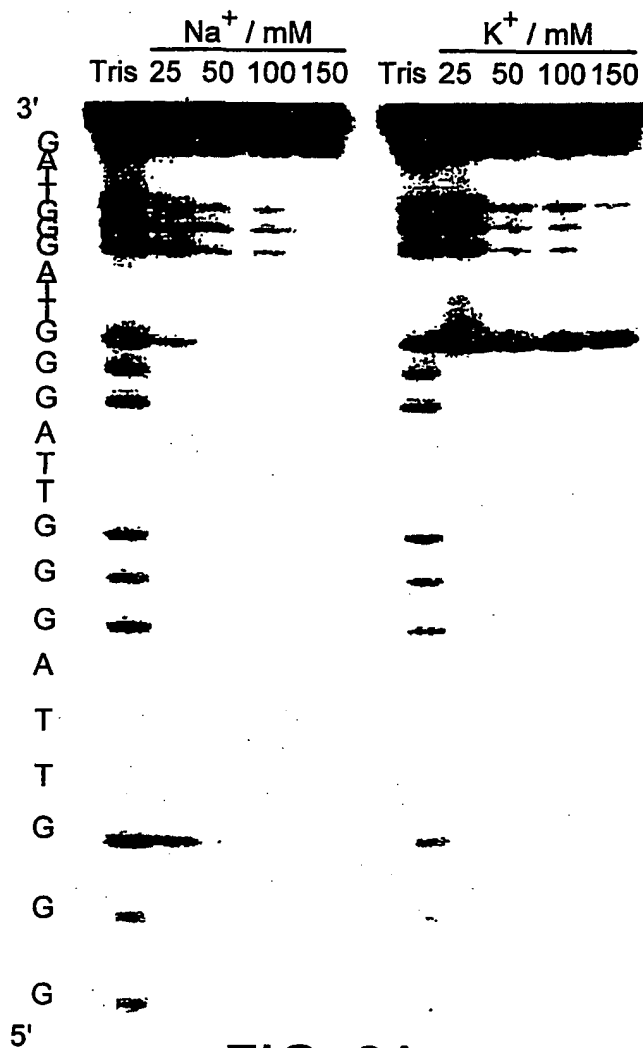


FIG. 2A

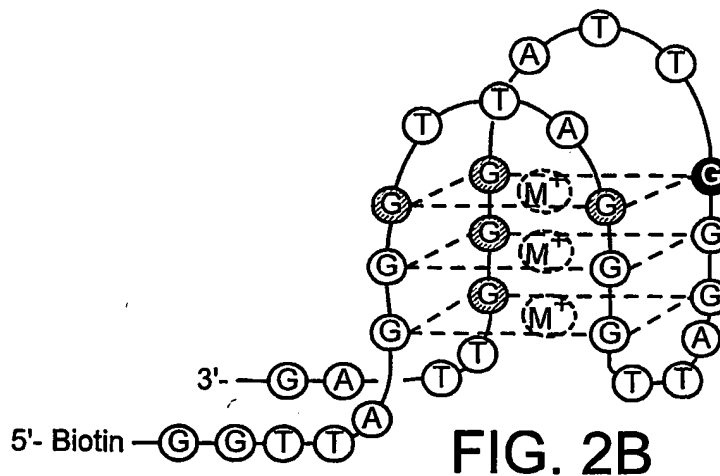


FIG. 2B

3 / 15

PROTEIN	F1							F2							F3									
	-1	1	2	3	4	5	6	-1	1	2	3	4	5	6	-1	1	2	3	4	5	6			
Gq1		D	S	A	H	L	T	R		D	R	S	D	L	S	E		R	S	D	H	R	I	E
Gq2		R	S	D	H	L	I	N		D	R	A	D	L	S	E		T	S	S	H	R	T	N
Gq3		D	S	A	H	L	T	R		D	R	D	H	L	S	E		T	S	S	H	R	T	N
Gq4		T	S	H	H	L	I	Q		D	R	A	D	L	S	E		H	Q	H	Y	R	T	N
Zif268		R	S	D	E	L	T	R		R	S	D	H	L	T	T		R	S	D	E	R	K	R

FIG. 3

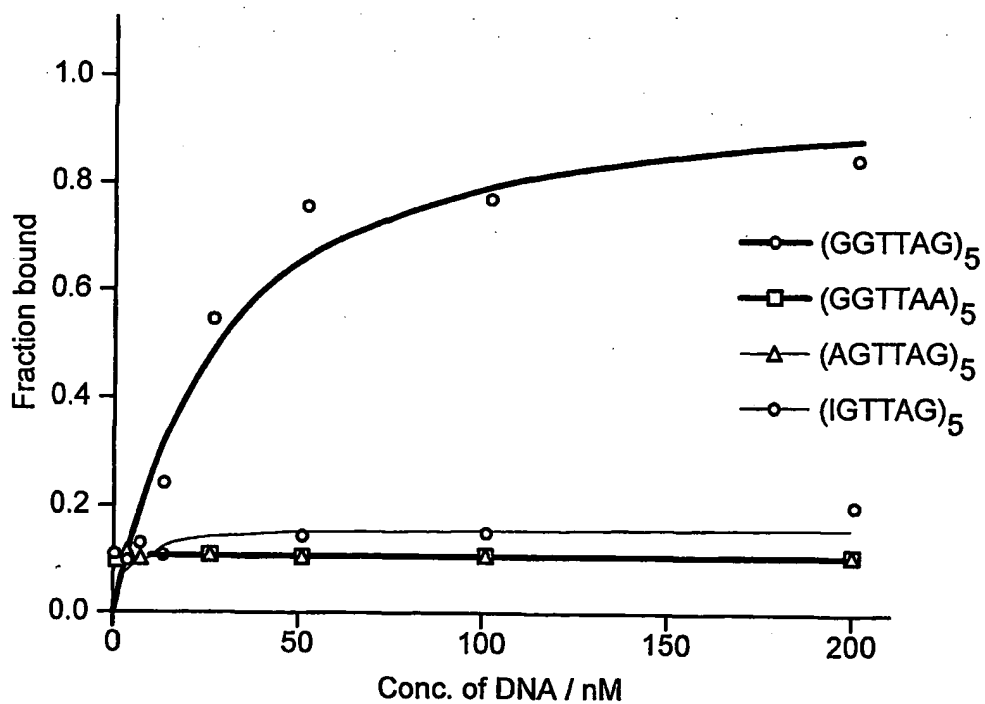


FIG. 4

4 / 15

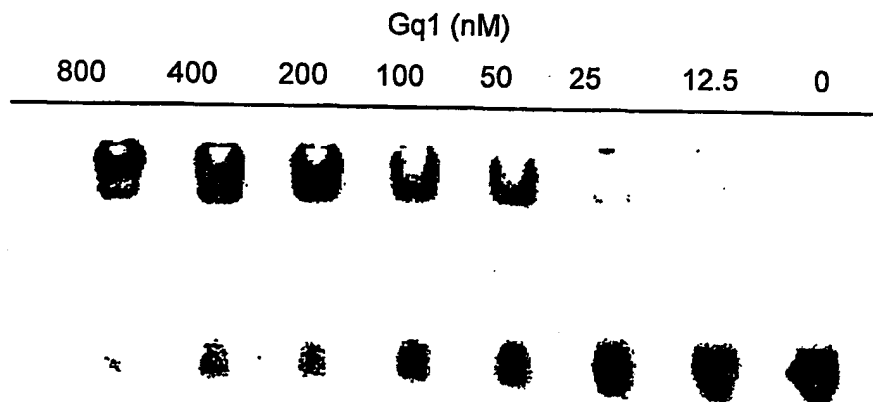


FIG. 5A

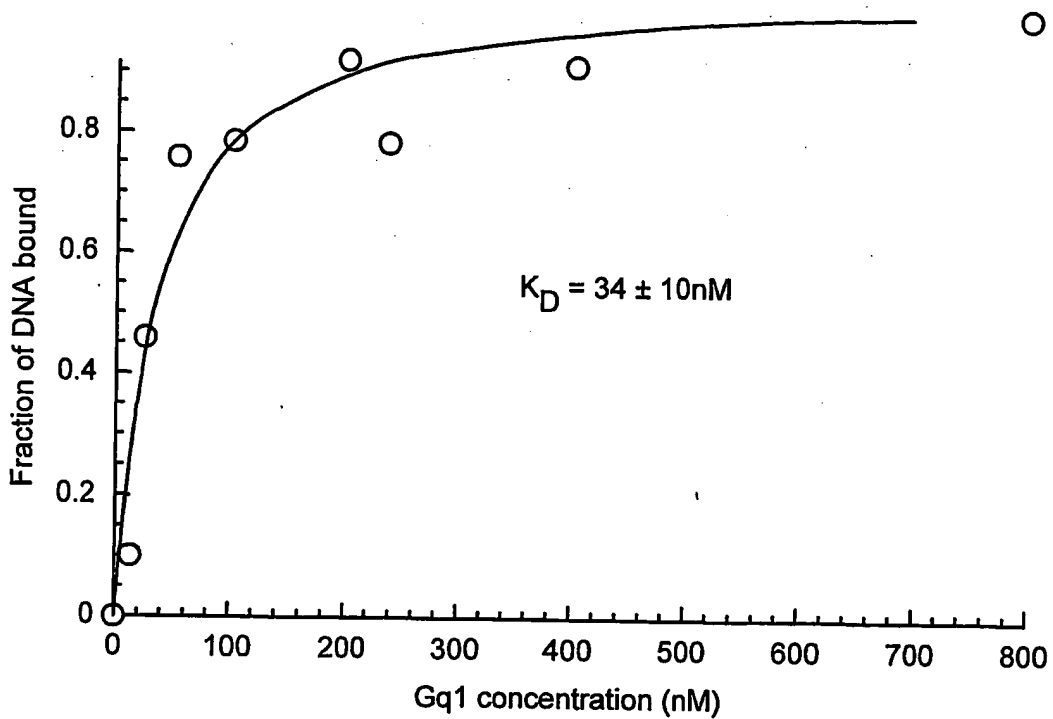


FIG. 5B

5 / 15

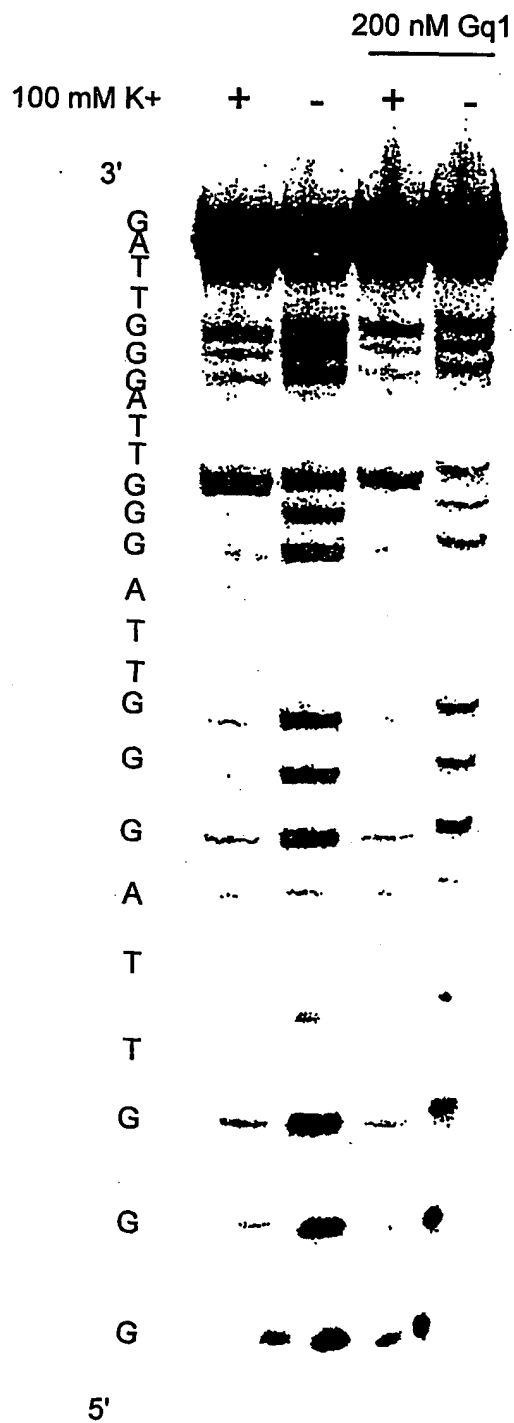


FIG. 6

6 / 15

Table 1

DNA sequence (5'-3')	K_d^E (nM)
(GGTTAG) ₅	26
(GGTTAA) ₅	-
(AGTTAG) ₅	-
(IGTTAG) ₅	-
(GGTTAG) ₅ duplex	-

FIG. 7

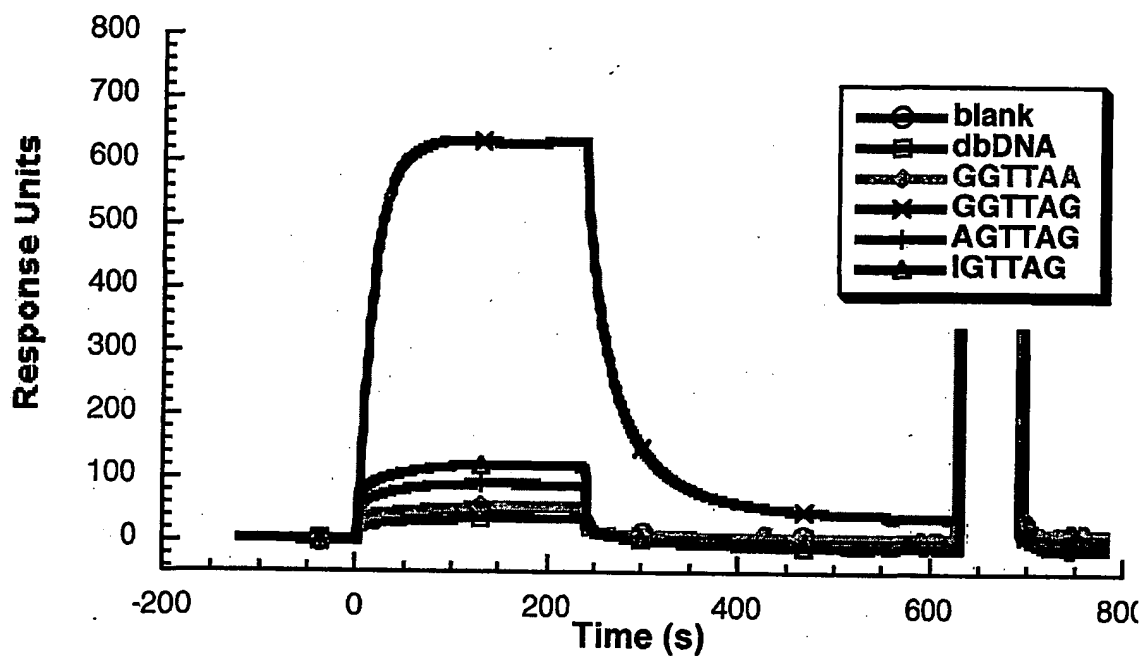


FIG. 8

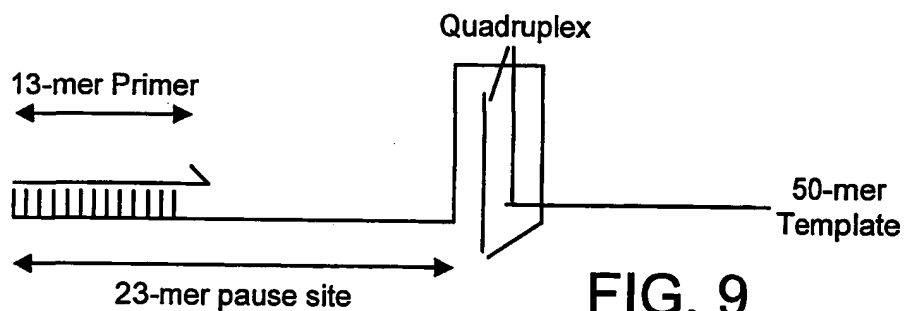


FIG. 9

7 / 15

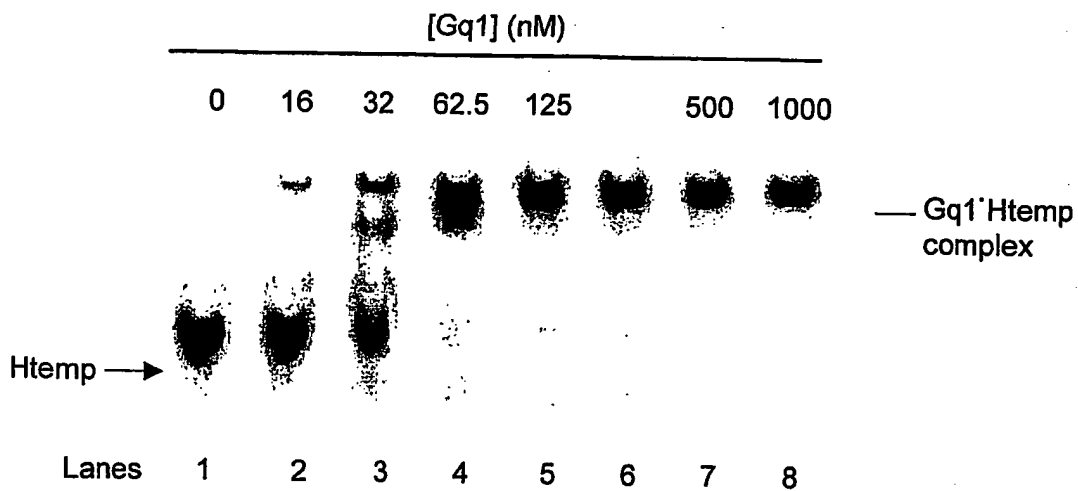


FIG. 10A

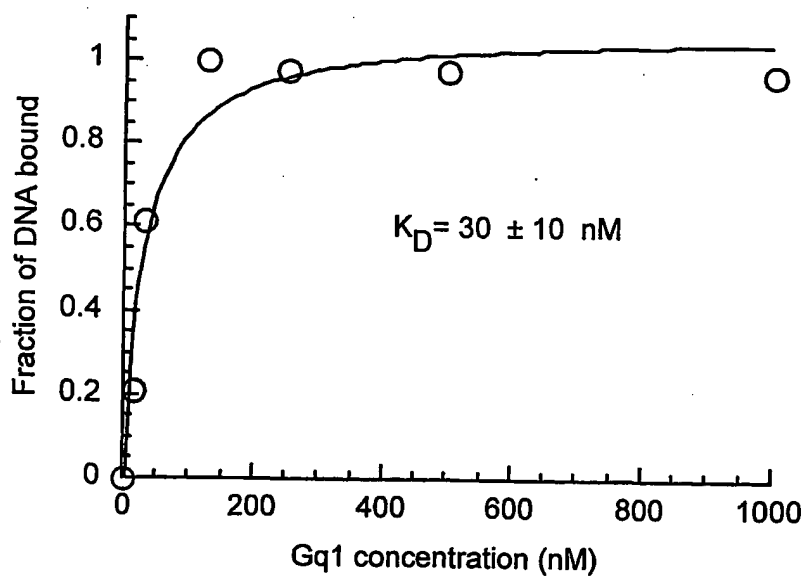


FIG. 10B

8 / 15

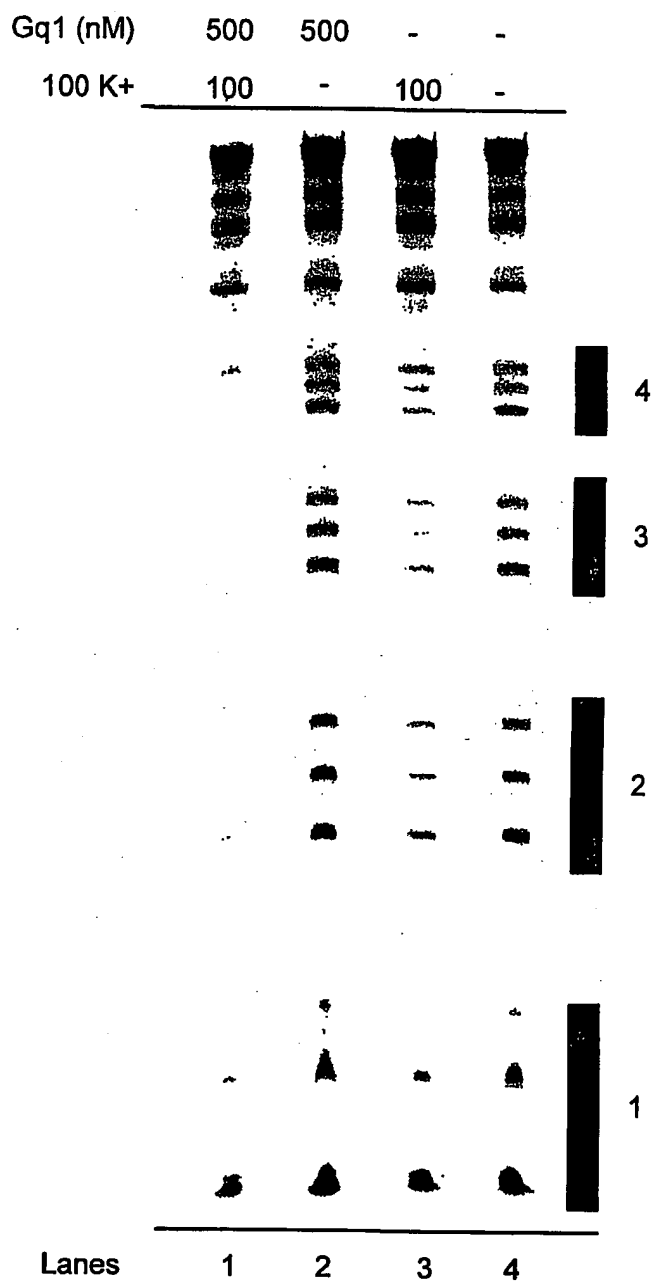


FIG. 11

9 / 15

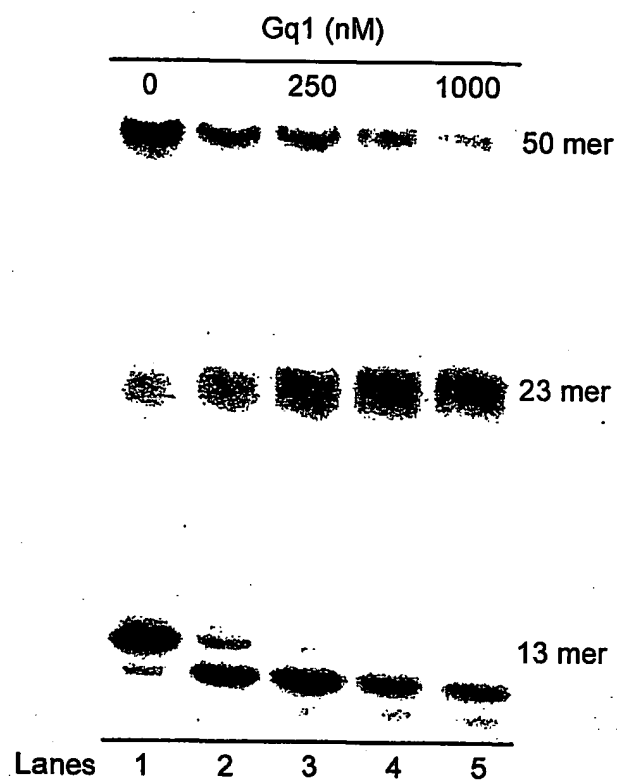


FIG. 12A

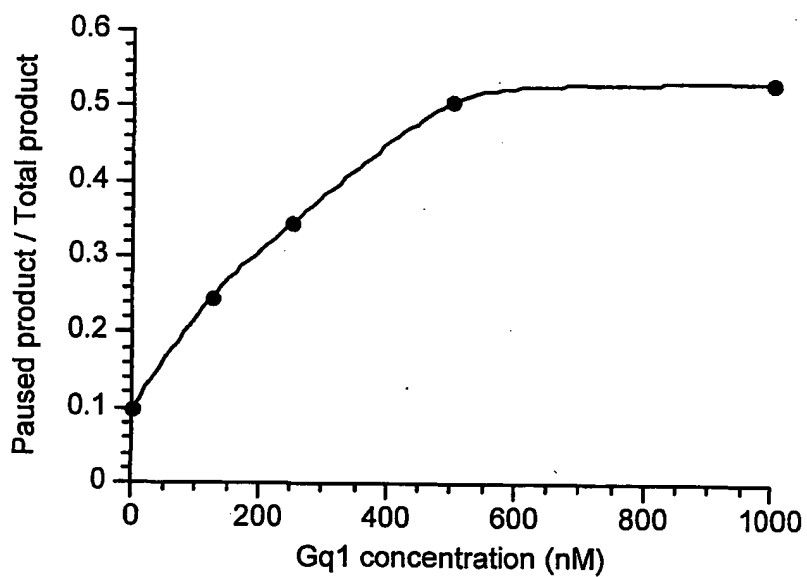


FIG. 12B

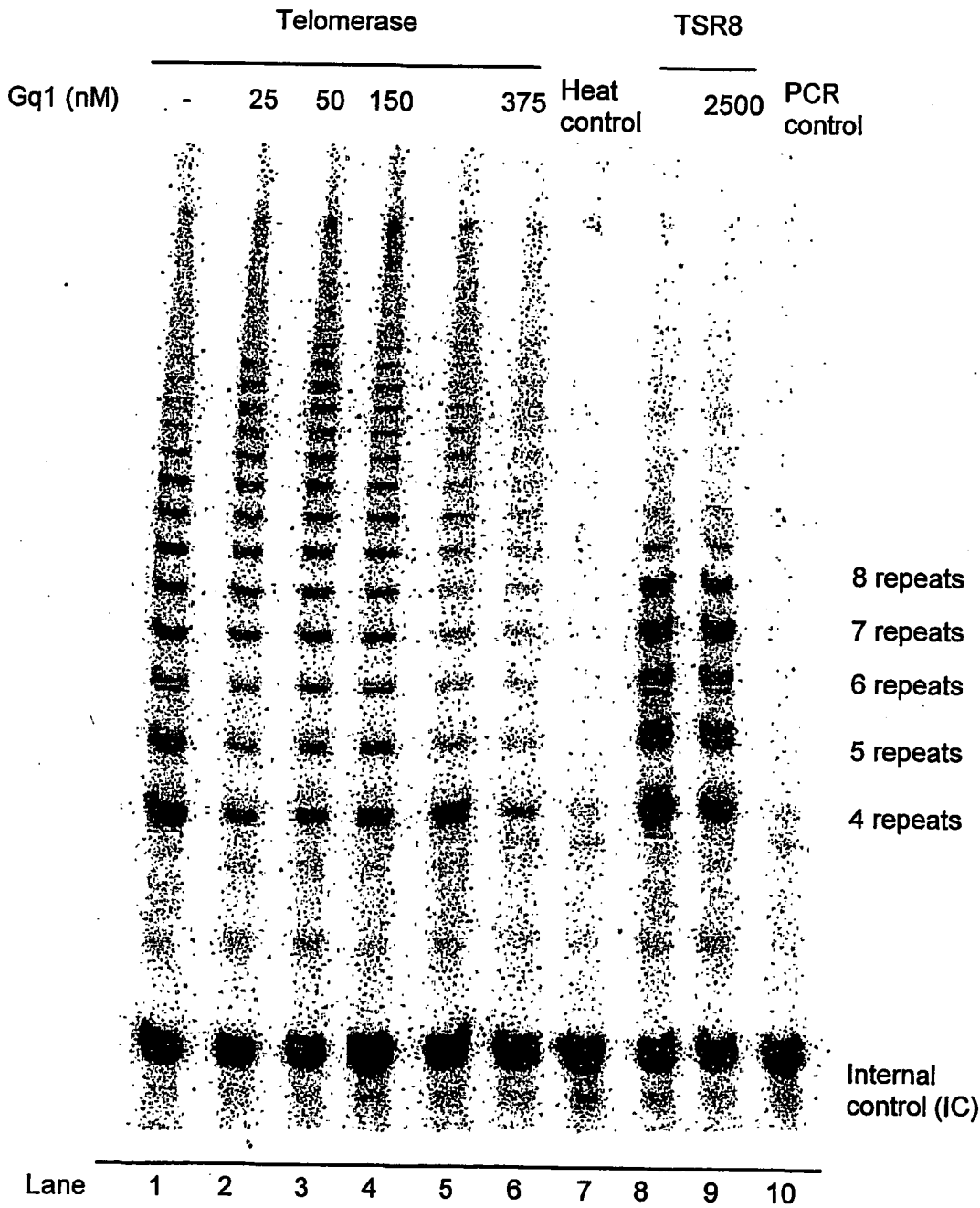


FIG. 13

11 / 15

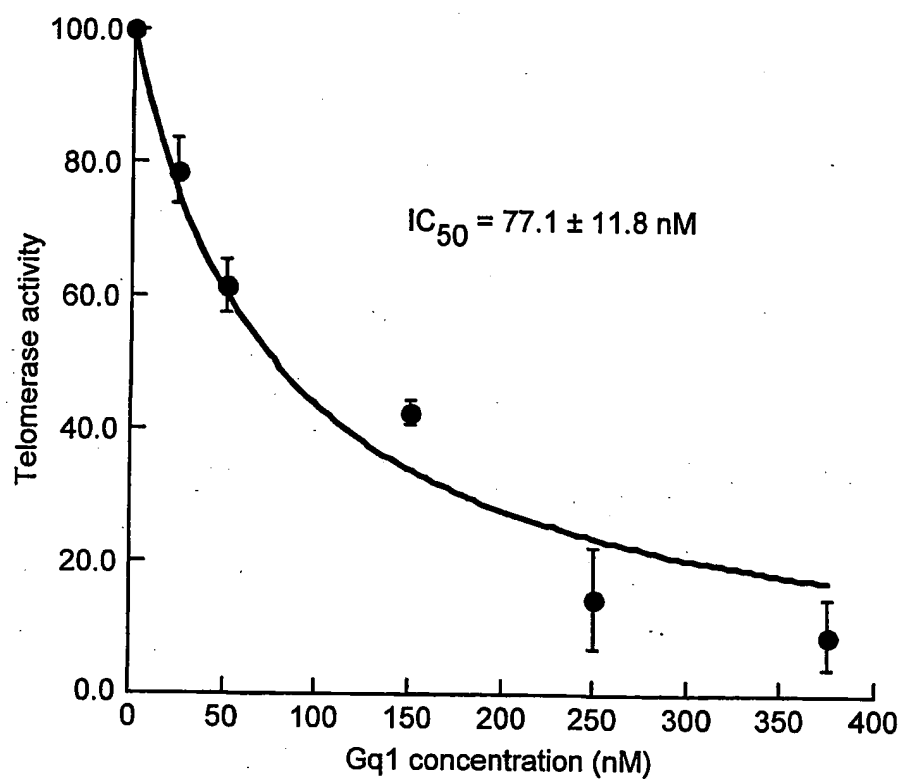


FIG. 14

12 / 15

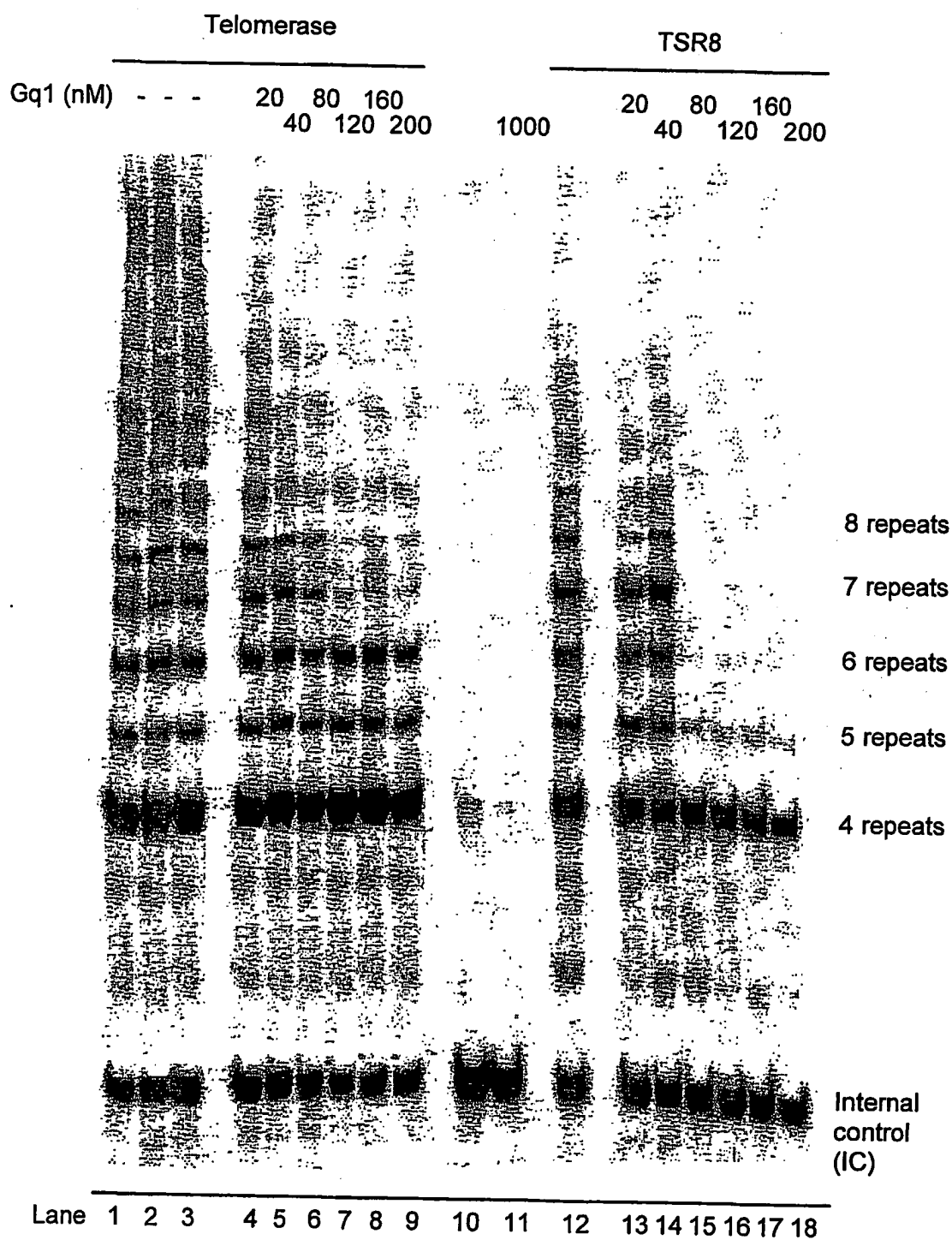


FIG. 15

13 / 15

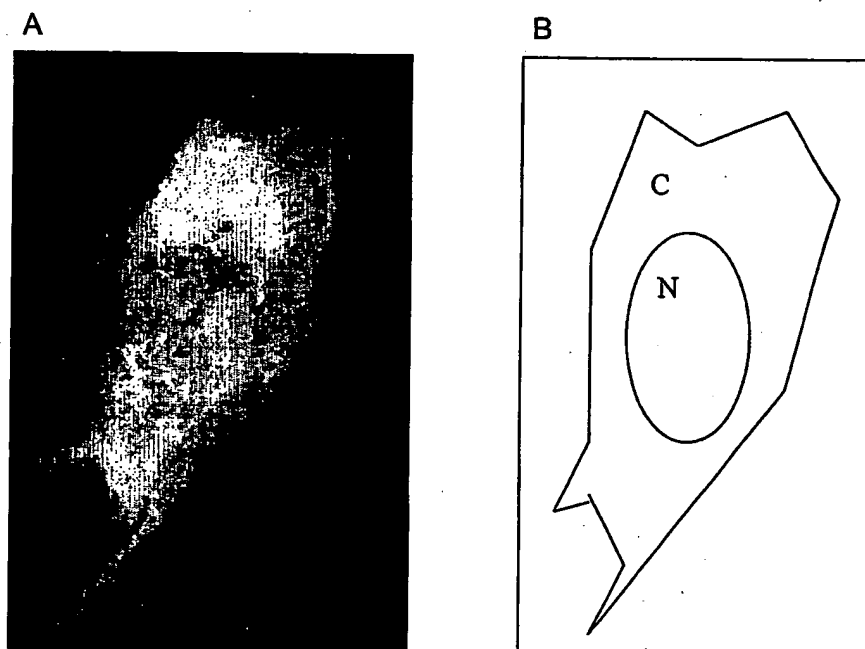


FIG. 16

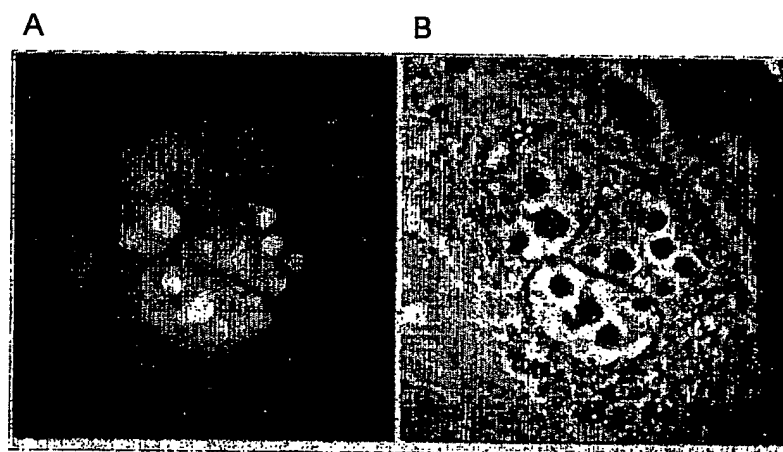


FIG. 17

14 / 15

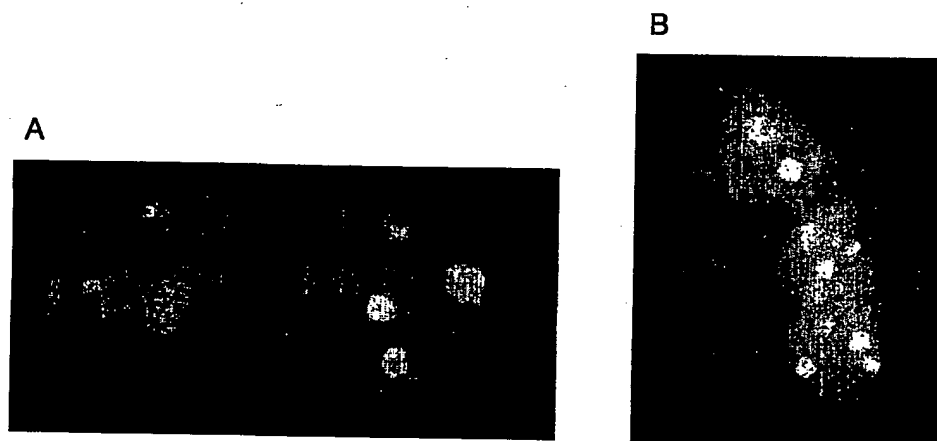


FIG. 18

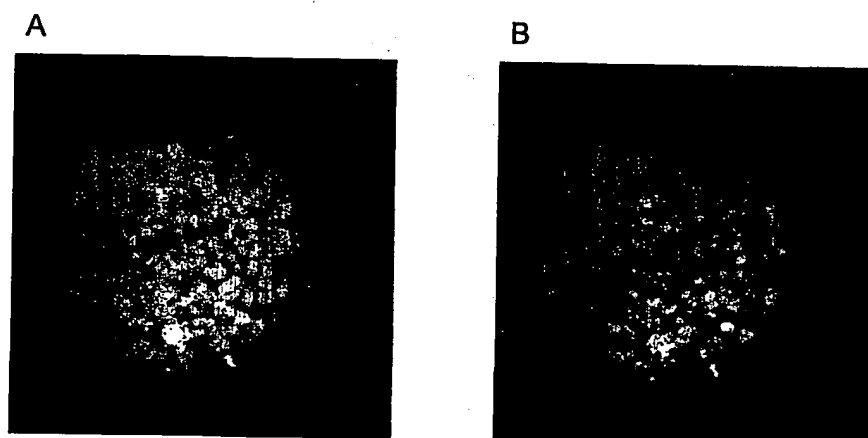


FIG. 19

15 / 15

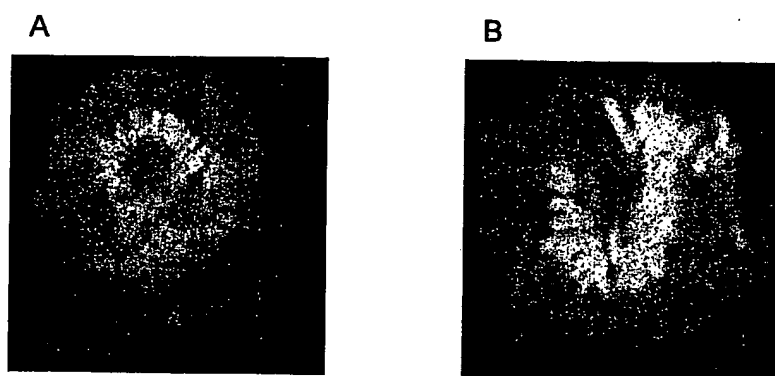


FIG. 20

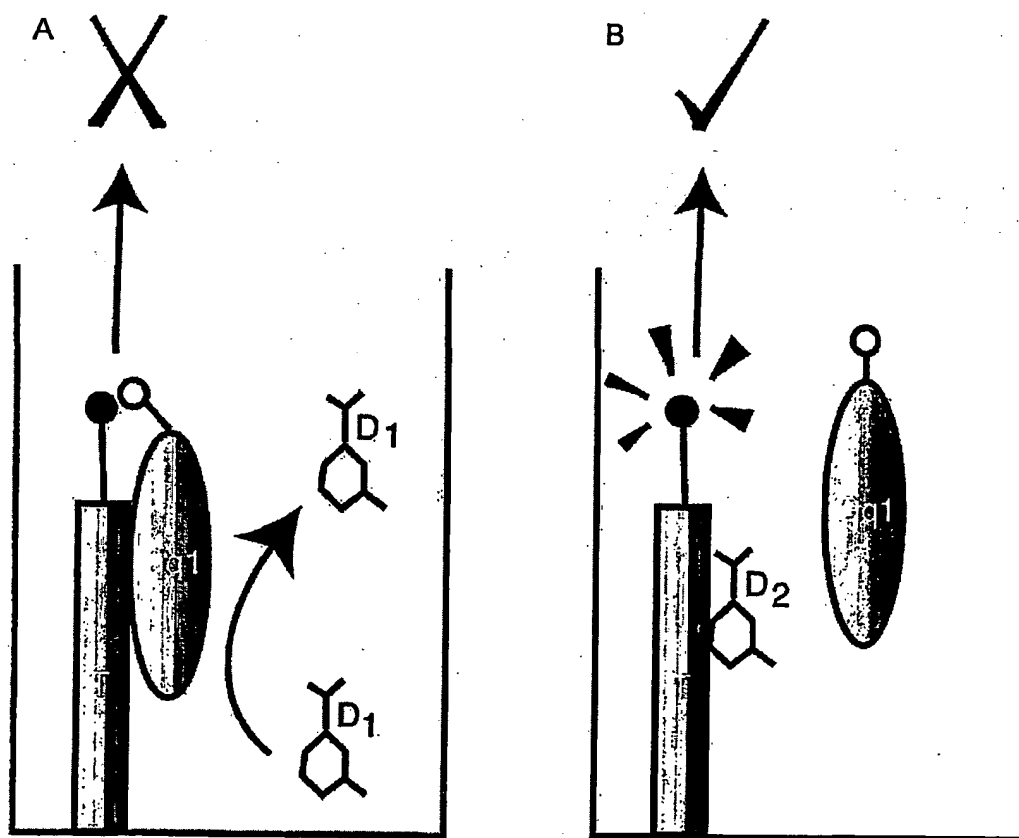


FIG. 21

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
17 January 2002 (17.01.2002)

PCT

(10) International Publication Number
WO 02/004488 A3

- (51) International Patent Classification⁷: C12N 15/12, C07K 14/47, G01N 33/68, A61K 38/00, 48/00
- (21) International Application Number: PCT/GB01/03130
- (22) International Filing Date: 12 July 2001 (12.07.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
09/614,679 12 July 2000 (12.07.2000) US
- (71) Applicants (for all designated States except US): SANG-AMO BIOSCIENCES, INC. [US/US]; Point Richmond Tech Center, Suite A100, 501 Canal Boulevard, Richmond, CA 94804 (US). CAMBRIDGE UNIVERSITY TECHNICAL SERVICES LTD. [GB/GB]; The Old Schools, Trinity Lane, Cambridge CB2 1TN (GB).

ISALAN, Mark [GB/GB]; 24 Shottfield Avenue, East Sheen, London SW14 8EA (GB). LIU, Xiaohai [CN/GB]; 136 Mowbray Road, Cambridge, CB1 7TG (GB). PATEL, Sachin [IN/IN]; Themis Medicare, 10/12 Udyog Nagar, S-V-RD, Goregon (W), Mumbai 400104 (IN). BALASUBRAMANIAN, Shankar [GB/GB]; Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW (GB).

(74) Agents: KHOO, Chong-Yee et al.; D Young & Co., 21 New Fetter Lane, London EC4A 1DA (GB).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

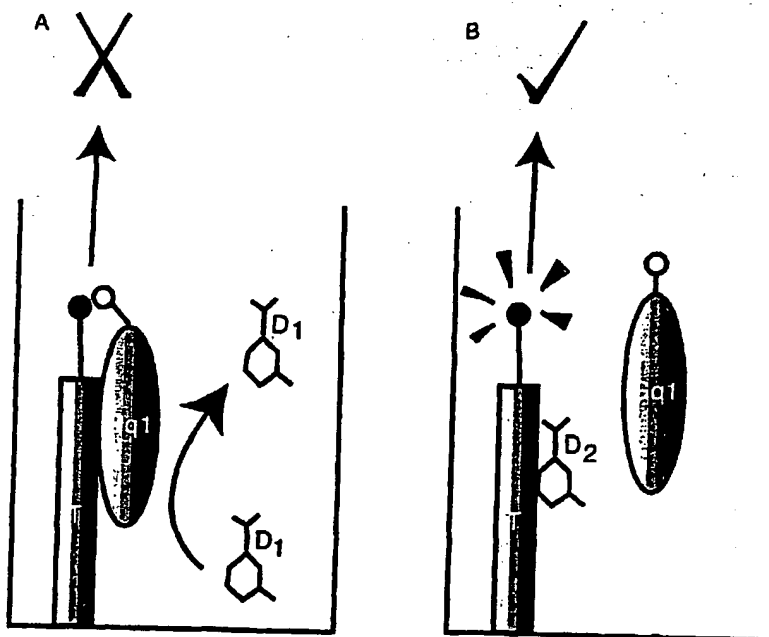
(72) Inventors; and

(75) Inventors/Applicants (for US only): CHOO, Yen [GR/GB]; 130 Ross Street, Cambridge CB1 3BU (GB).

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian

[Continued on next page]

(54) Title: NUCLEIC ACID BINDING POLYPEPTIDES



(57) Abstract: We describe the use of a nucleic acid binding polypeptide capable of binding to one or more of telomeric, G-quadruplex, or G-quartet nucleic acid as an inhibitor of enzymatic activity, including a telomerase activity, a polymerase activity, an integrase activity and a gp120 activity. Telomerase assays as well as methods of identifying molecules capable of interacting with telomeric, G-quadruplex, or G-quartet nucleic acid are also described.



patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(88) Date of publication of the international search report:
18 July 2002

Declaration under Rule 4.17:

— of inventorship (Rule 4.17(iv)) for US only

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

— with international search report

INTERNATIONAL SEARCH REPORT

Inventor's Application No

PCT/GB 01/03130

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C07K14/47 G01N33/68 A61K38/00 A61K48/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K C12Q G01N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

WPI Data, PAJ, CAB Data, SEQUENCE SEARCH, BIOSIS, EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	ISALAN MARK ET AL: "Selection of zinc fingers that bind single-stranded telomeric DNA in the G-quadruplex conformation." BIOCHEMISTRY, vol. 40, no. 3, 23 January 2001 (2001-01-23), pages 830-836, XP002195928 ISSN: 0006-2960 the whole document — —/—	1-46

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "A" document member of the same patent family

Date of the actual completion of the international search

15 April 2002

Date of mailing of the international search report

29/04/2002

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Hornig, H

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 01/03130

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MAZUMDER ABHIJIT ET AL: "Inhibition of the human immunodeficiency virus type 1 integrase by guanosine quartet structures" BIOCHEMISTRY, AMERICAN CHEMICAL SOCIETY, EASTON, PA, US, vol. 35, no. 43, 29 October 1996 (1996-10-29), pages 13762-13771, XP002148822 ISSN: 0006-2960 the whole document	
A	ZAHLER A M ET AL: "INHIBITION OF TELOMERASE BY G-QUARTET DNA STRUCTURES" NATURE, MACMILLAN JOURNALS LTD. LONDON, GB, vol. 350, 25 April 1991 (1991-04-25), pages 718-720, XP002195929 ISSN: 0028-0836 the whole document	
A	WO 98 53057 A (MEDICAL RES COUNCIL ;ISALAN MARK (GB); CHOO YEN (GB); KLUG AARON () 26 November 1998 (1998-11-26) cited in the application the whole document	
A	WO 96 06166 A (MEDICAL RES COUNCIL ;CHOO YEN (SG); KLUG AARON (GB); GARCIA ISIDRO) 29 February 1996 (1996-02-29) the whole document	
A	WO 98 33503 A (WHEELHOUSE RICHARD T ;HURLEY LAURENCE H (US); UNIV TEXAS (US)) 6 August 1998 (1998-08-06) the whole document	
A	US 6 001 657 A (ROBERTS JOHN F ET AL) 14 December 1999 (1999-12-14) the whole document	
A	J.C. BROWN ET AL.: "Construction and characterization of a quadruplex DNA selective single-chain autoantibody from a viable motheaten mouse hybridoma with homology to telomeric DNA binding proteins" BIOCHEMISTRY, vol. 37, no. 46, March 1998 (1998-03), pages 16338-16348, XP002195930 AM. CHEM. SOC., WASHINGTON, DC, US the whole document	

-/--

INTERNATIONAL SEARCH REPORT

Int'l Application No
PCT/GB 01/03130

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>B.A. BROWN II ET AL.: "Isolation and characterization of a monoclonal anti-quadruplex DNA antibody from autoimmune "viable motheaten" mice" BIOCHEMISTRY, vol. 37, no. 46, March 1998 (1998-03), pages 16325-16337, XP002195931 AM. CHEM. SOC., WASHINGTON, DC, US cited in the application the whole document</p> <p>-----</p>	

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PLT/GB 01/03130

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 9853057	A	26-11-1998	AU 737756 B2	30-08-2001
			AU 7542298 A	11-12-1998
			AU 7542398 A	11-12-1998
			AU 732017 B2	12-04-2001
			AU 7542698 A	11-12-1998
			EP 0983349 A1	08-03-2000
			EP 0983350 A1	08-03-2000
			EP 0983351 A1	08-03-2000
			WO 9853057 A1	26-11-1998
			WO 9853058 A1	26-11-1998
			WO 9853060 A1	26-11-1998
			JP 2002502238 T	22-01-2002
			JP 2001527417 T	25-12-2001
WO 9606166	A	29-02-1996	AU 1003799 A	22-04-1999
			AU 698152 B2	22-10-1998
			AU 3229195 A	14-03-1996
			CA 2196419 A1	29-02-1996
			EP 0781331 A1	02-07-1997
			WO 9606166 A1	29-02-1996
			JP 10504461 T	06-05-1998
			US 6013453 A	11-01-2000
WO 9833503	A	06-08-1998	US 6007988 A	28-12-1999
			AU 6650198 A	25-08-1998
			EP 0988037 A1	29-03-2000
			US 6087493 A	11-07-2000
US 6001657	A	14-12-1999	WO 9833503 A1	06-08-1998
			NONE	

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER: _____**

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.